Challenges in Modeling for Engineering/Deciphering Regulatory Networks

James C. Liao
Department of Chemical and Biomolecular Engineering
University of California, Los Angeles

Modeling

- What is the purpose of the model?
- What data to use?
- What are the existing/alternative approaches?
- What is the scope of the model?
- Mathematical details?
- What is the basis of prediction?

Modeling

- What is the purpose of the model?
  - To elucidate active Nitric Oxide (NO) response networks in E. coli
- What data to use?
- What are the existing/alternative approaches?
- What is the scope of the model?
- Mathematical details?
- What is the basis of prediction?

E. coli, Nitric Oxide, and Septic Shock

NO is a mammalian signaling molecule

- vasodilator
- neurotransmitter
- smooth muscle relaxant
- **immune response**

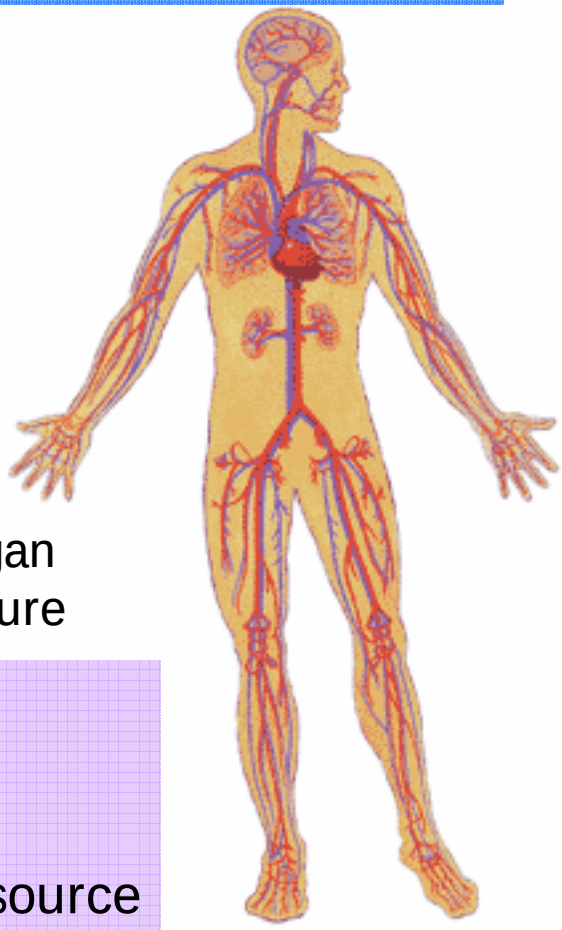
septicshock.com

Septic shock

NO production → vasodilation → hypotension → organ failure

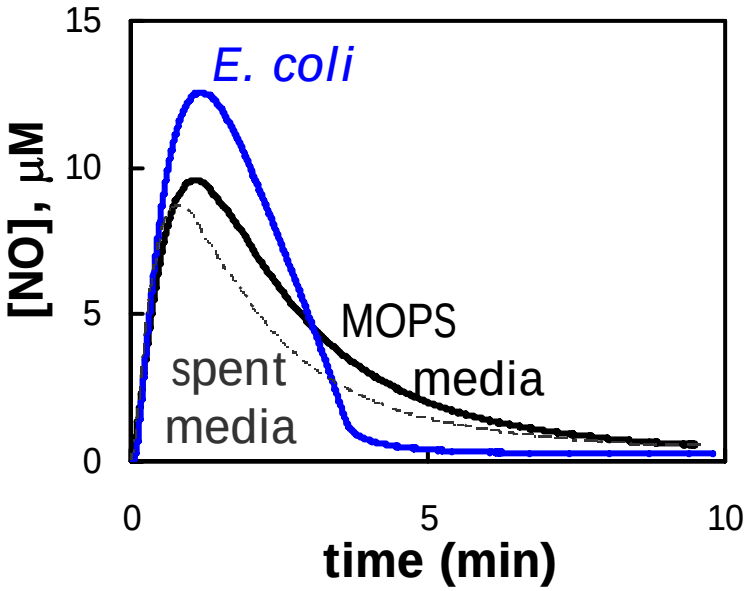
- average patient cost $50,000
- 10th leading cause of death (2000)

- urinary tract (UTI) major infection source
- *E. coli* accounts for ~90% of UTIs

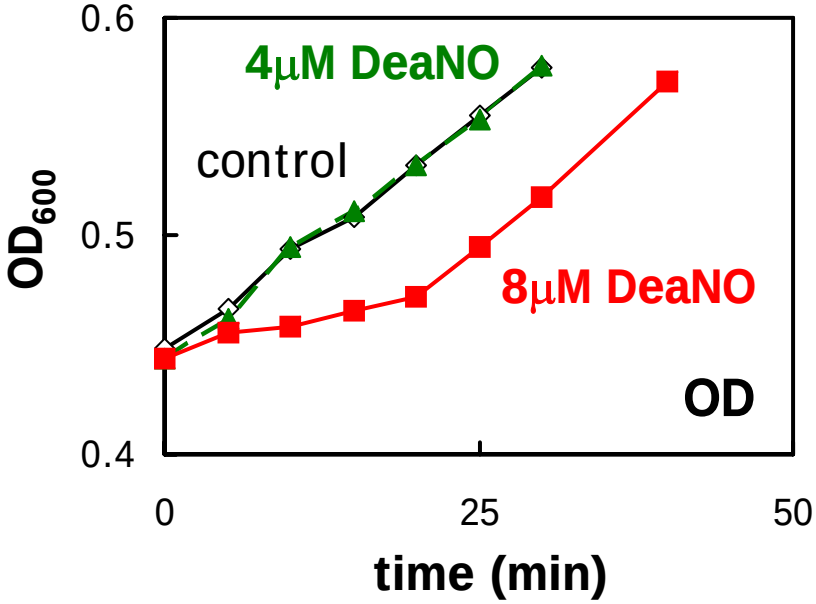


E. coli NO Response

NO Release from 10μM DeaNO



***E. coli* Growth Response**



- NO consumption increases in the presence of *E. coli*

DeaNO inhibits growth for a period longer than the NO lifetime

What causes the growth inhibition?

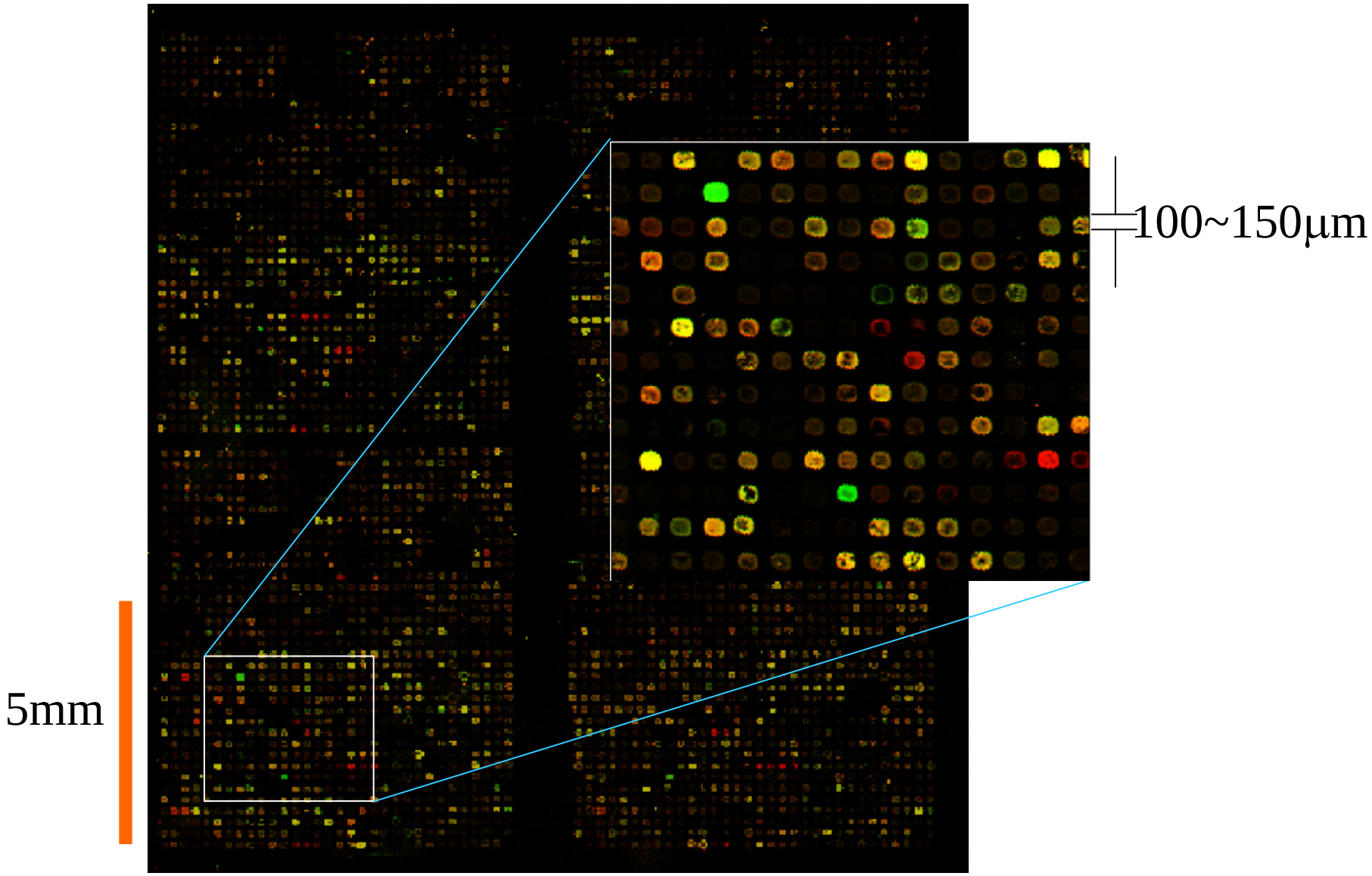
DeaNO = *diethylamine NONOate*

$t_{1/2}$ *(37^{0}C, pH 7.4) = 2 min*

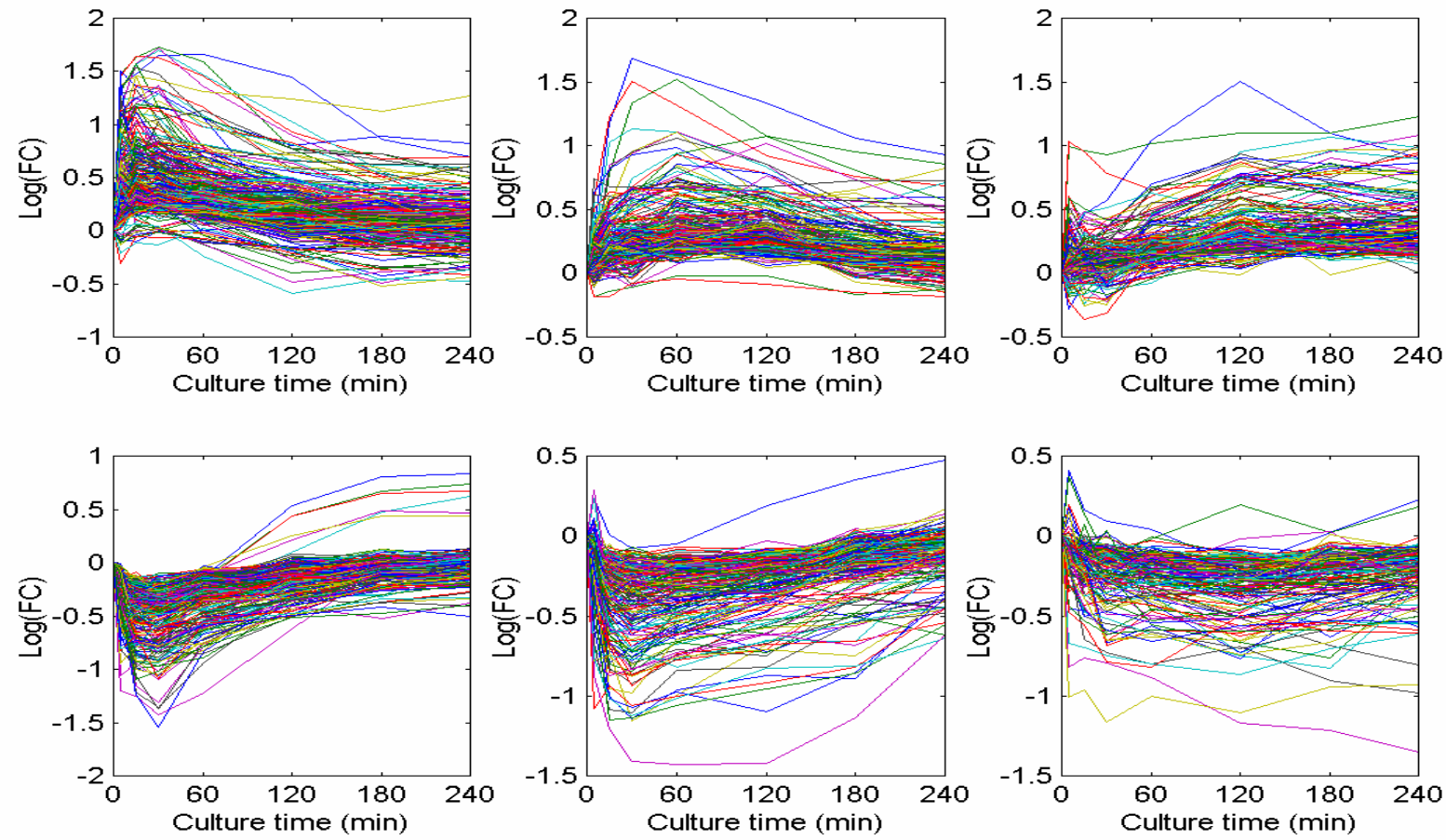
Modeling

- What is the purpose of the model?
  - To elucidate active Nitric Oxide (NO) response networks in E. coli
- What data to use?
  - Microarray
- What are the existing/alternative approaches?
- What is the scope of the model?
- Mathematical details?
- What is the basis of prediction?

E. coli whole-genome Microarray



Microarray Data



What do we learn from these data?

Modeling

- What is the purpose of the model?
  - To elucidate the NO response network.
- What are the existing/alternative approaches?
  - ad hoc analysis, clustering…
- What is the scope of the model?
- What data to use?
- Mathematical details?
- What is the basis of prediction?

Traditional Microarray Analysis Looks at One Gene at a Time

in WT

Gene Name	Fold Change
carA	-4.76
sdhA	-4.72
sdhC	-4.72
sdhB	-3.91
ndk	-3.72
sdhD	-3.46
argC	-3.03
metF	-3.03
yeeF	-2.90
b0725	-2.89
iscU	3.13
iscS	3.15
iscR	3.93
cydB	4.34
b1797	4.95
cydA	5.48
b4209	16.88

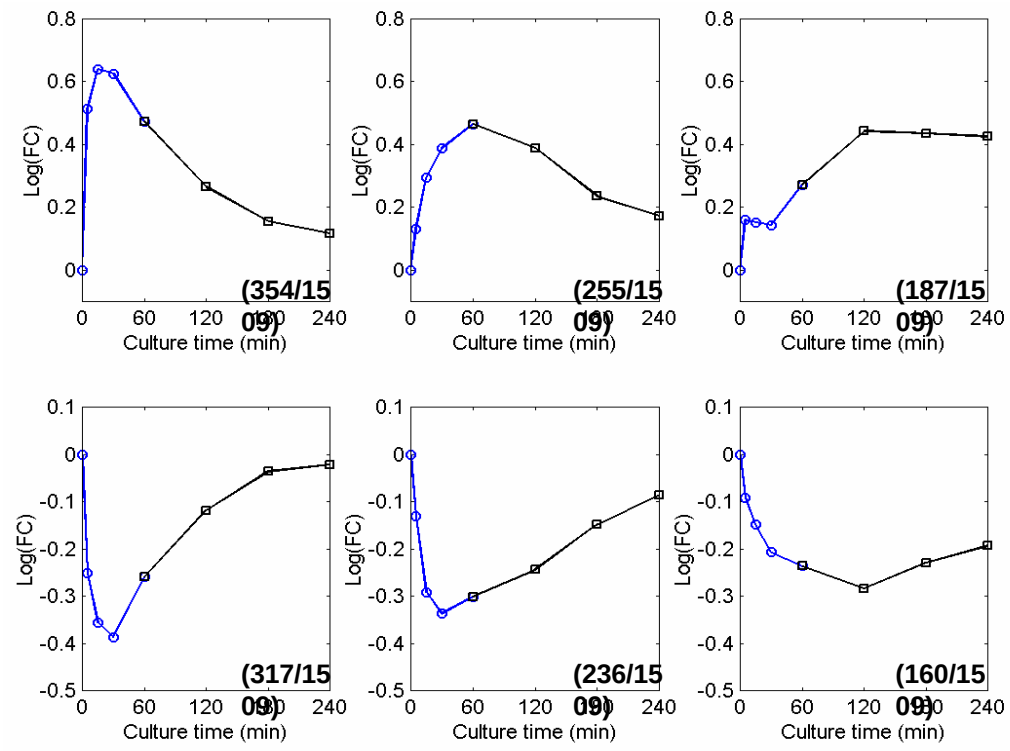
in ΔiscR

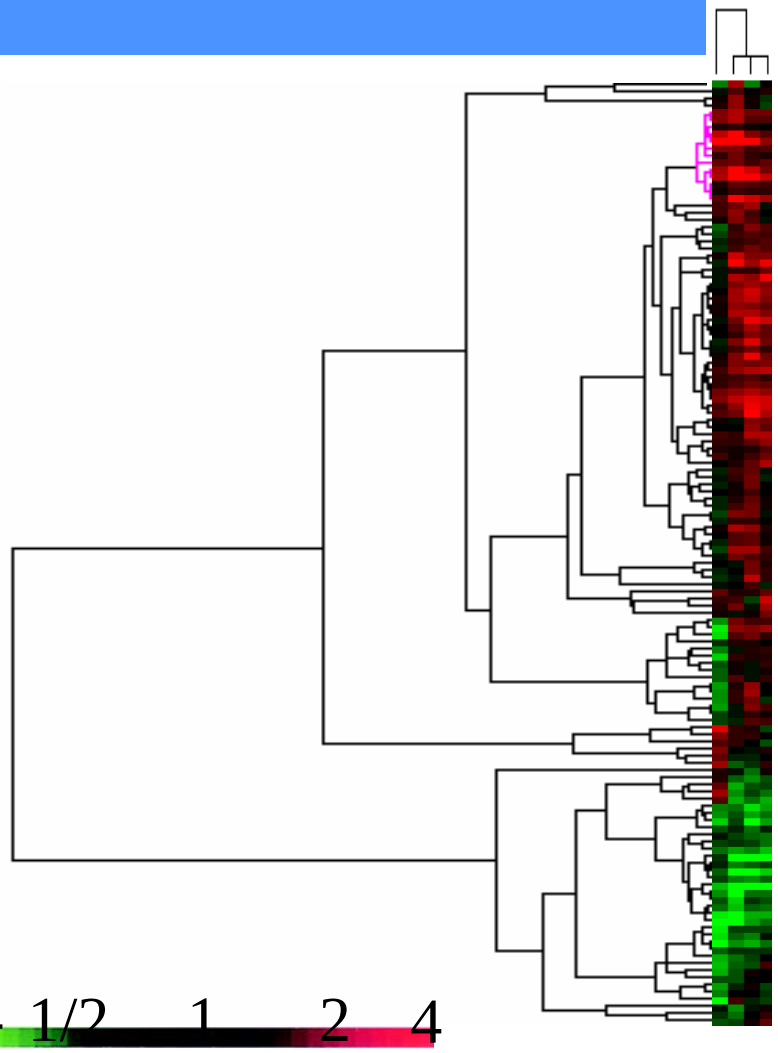
Gene Name	Fold Change
b0725	-3.16
sdhA	-3.04
sucB	-2.79
sdhC	-2.69
sdhD	-2.19
nadD	-2.17
sucA	-2.09
fdoI	-1.96
sdhB	-1.90
b2984	-1.86
yjgV	1.79
cydA	1.89
ompT	1.98
cydB	2.57
b1797	3.05
oppA	3.22
norW	5.67

in ΔarcA

Gene Name	Fold Change
carA	-10.22
cvpA	-3.76
metF	-3.28
yeeF	-3.19
purM	-2.96
rnpA	-2.81
ndk	-2.80
purH	-2.71
pyrD	-2.66
codB	-2.52
iscS	3.50
yhiO	3.95
iscR	4.08
ilvC	4.20
b0939	5.62
b1797	5.91
norV	15.69

Cluster Analysis





Modeling

- What is the purpose of the model?
  - To elucidate the NO response network.
- What are the existing/alternative approaches?
  - ad hoc analysis, clustering…
- What is the scope of the model?
  - mRNA, transcription factors, degradation factors…
- What data to use?
- Mathematical details?
- What is the basis of prediction?

Modeling

- What is the purpose of the model?
  - To elucidate the NO response network.
- What are the existing/alternative approaches?
  - ad hoc analysis, clustering…
- What is the scope of the model?
  - mRNA, transcription factors, degradation factors…
- What data to use?
  - Microarray transcriptome data
- Mathematical details?
- What is the basis of prediction?

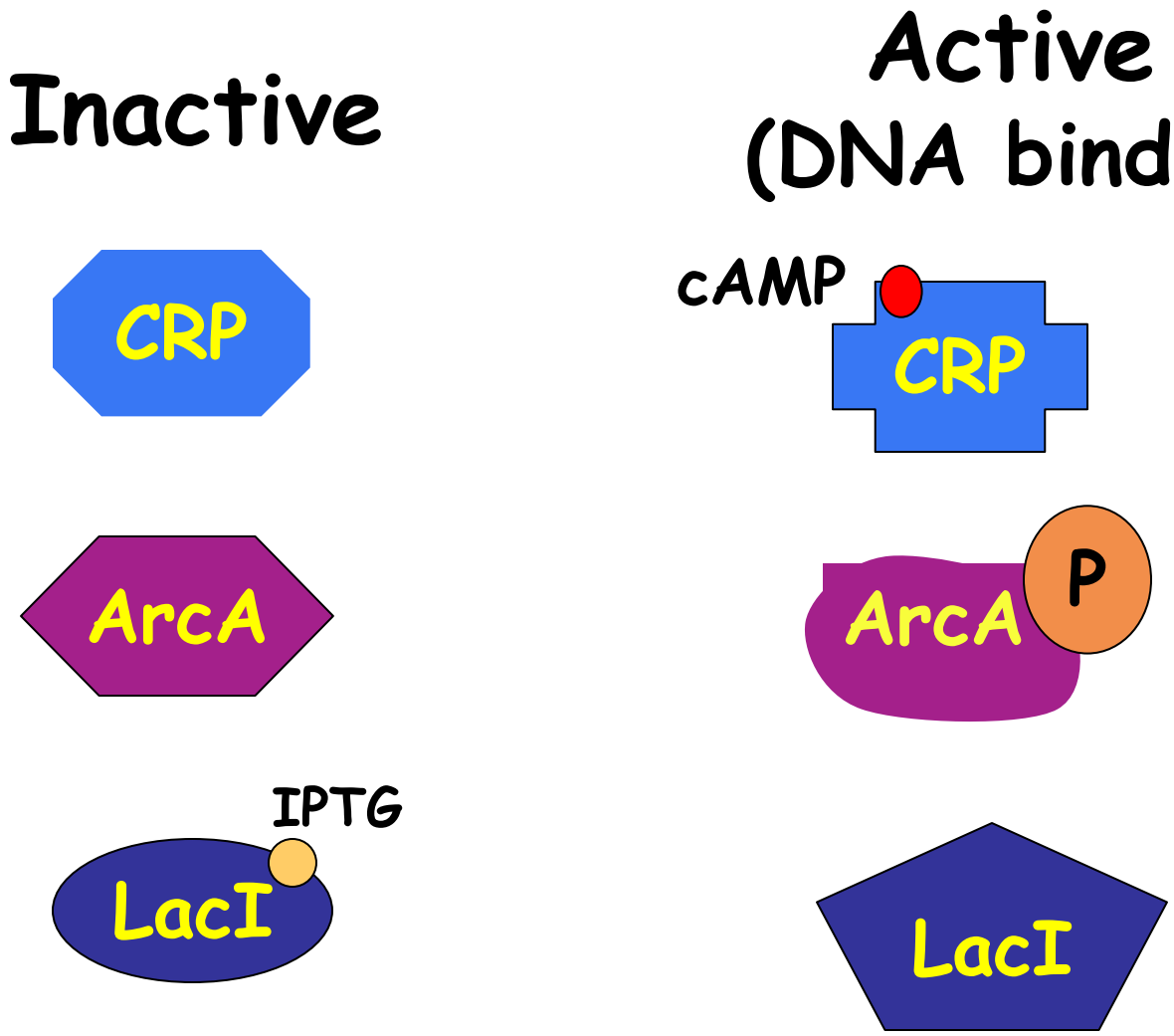
Characteristics of Transcription Networks

- Trascription factors regulate transcription
- Transcription factors are regulated post-transcriptionally
- Active transcription factor binds to transcription complex.
- Effect of binding is gene-dependent
- Binding is condition-dependent

Characteristics of Transcription Networks

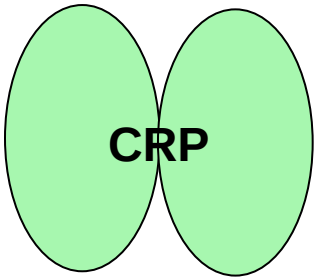
- Trascription factors regulate transcription
- Transcription factors are regulated post-transcriptionally
- Active transcription factor binds to transcription complex.
- Effect of binding is gene-dependent
- Binding is condition-dependent

Transcription Factor Activity (TFA)



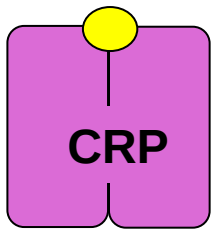
CRP induction mechanism

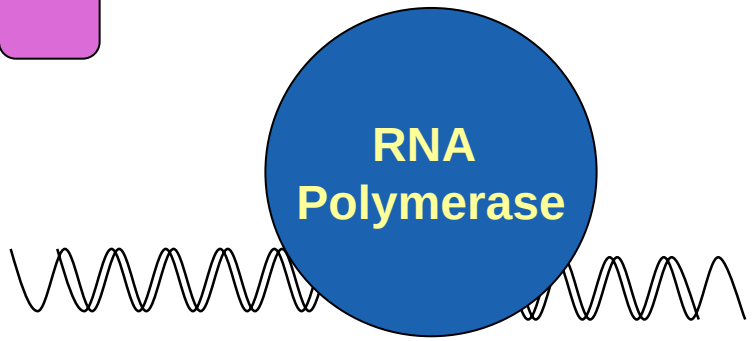
CRP (cAMP Receptor Protein)



Inactive

cAMP



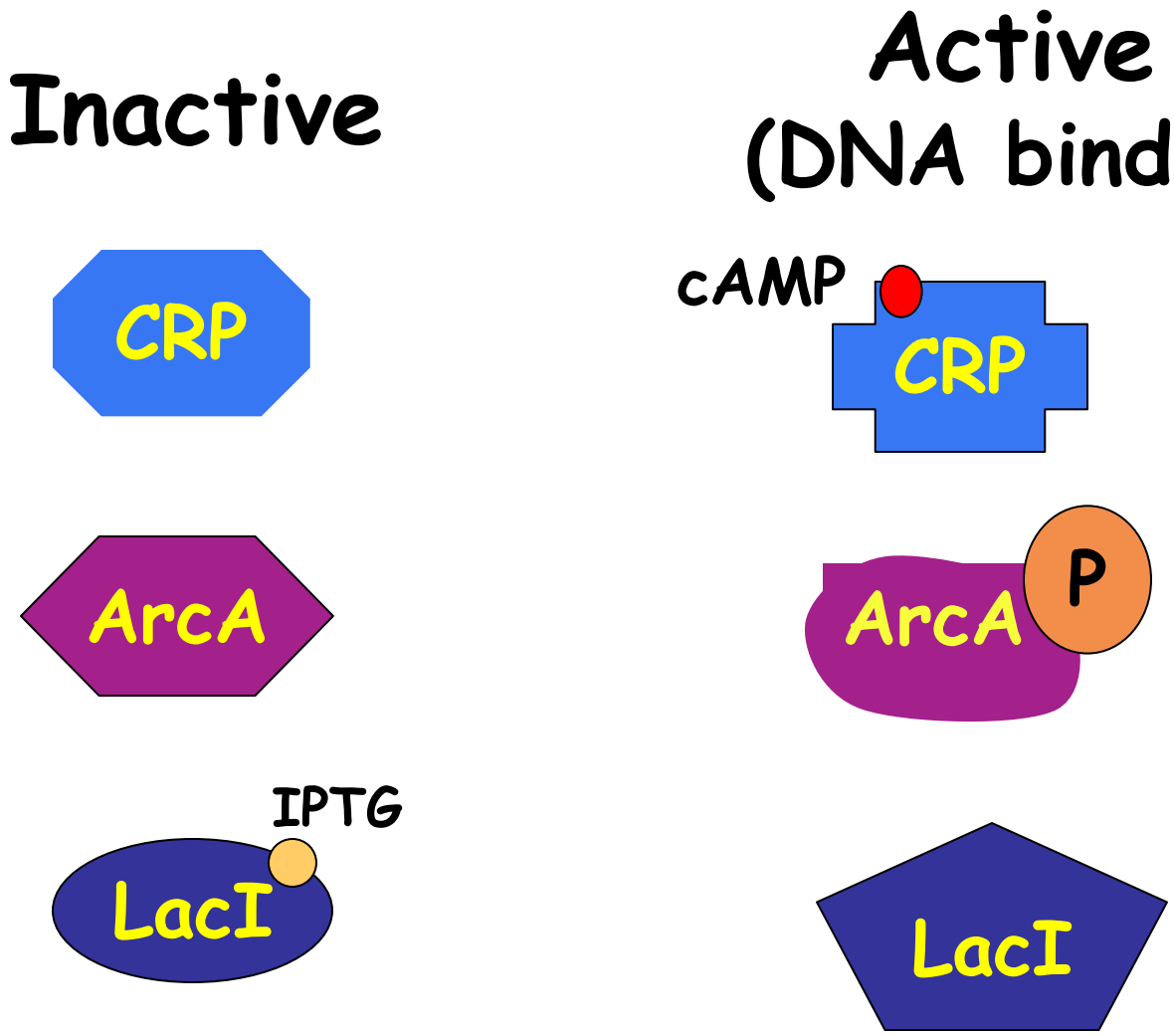


Active

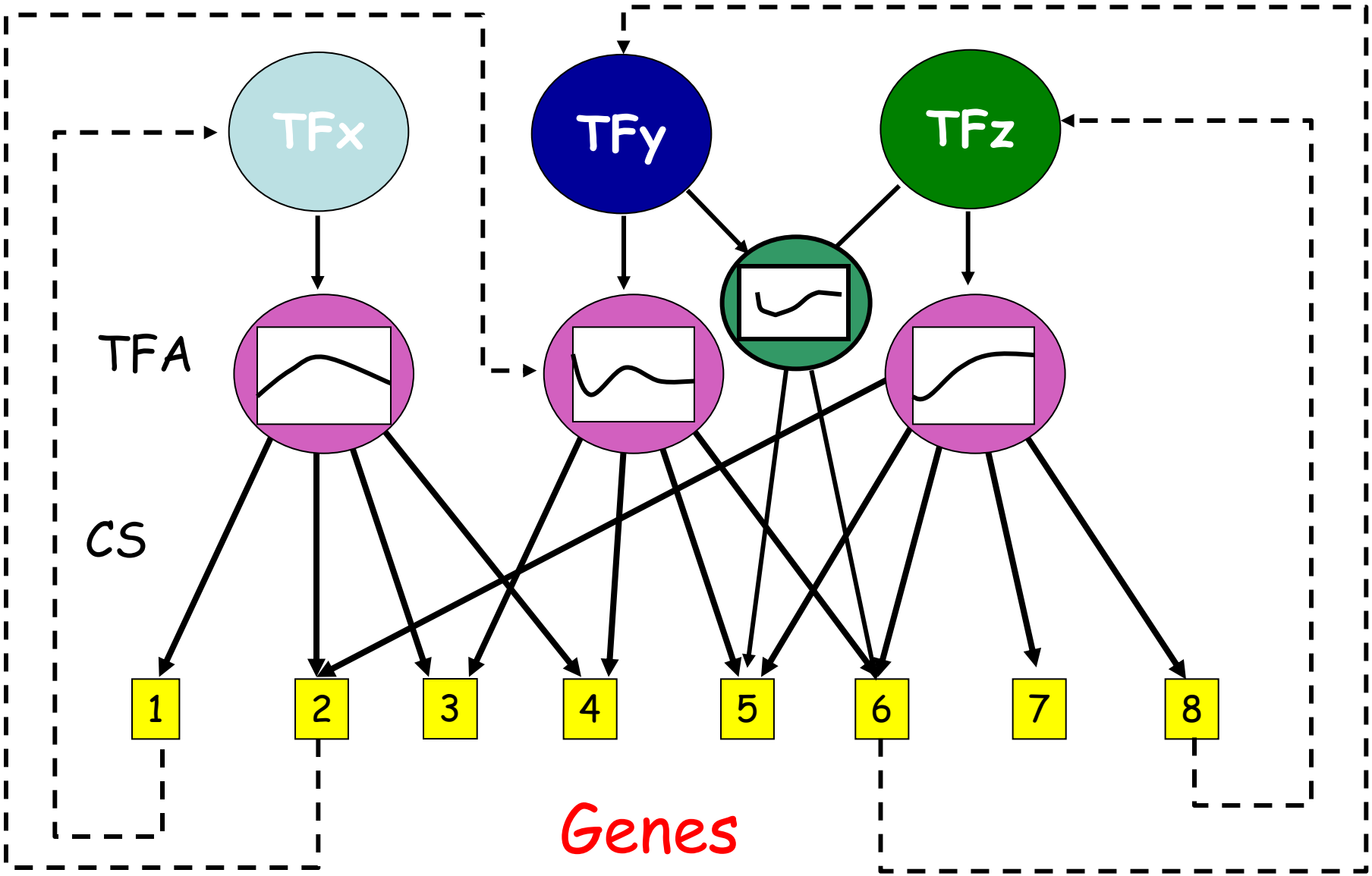
Characteristics of Transcription Networks

- Trascription factors regulate transcription
- Transcription factors are regulated post-transcriptionally
- Active transcription factors bind to transcription complex.
- Effect of binding is gene-dependent
- Binding is condition-dependent

Transcription Factor Activity (TFA)



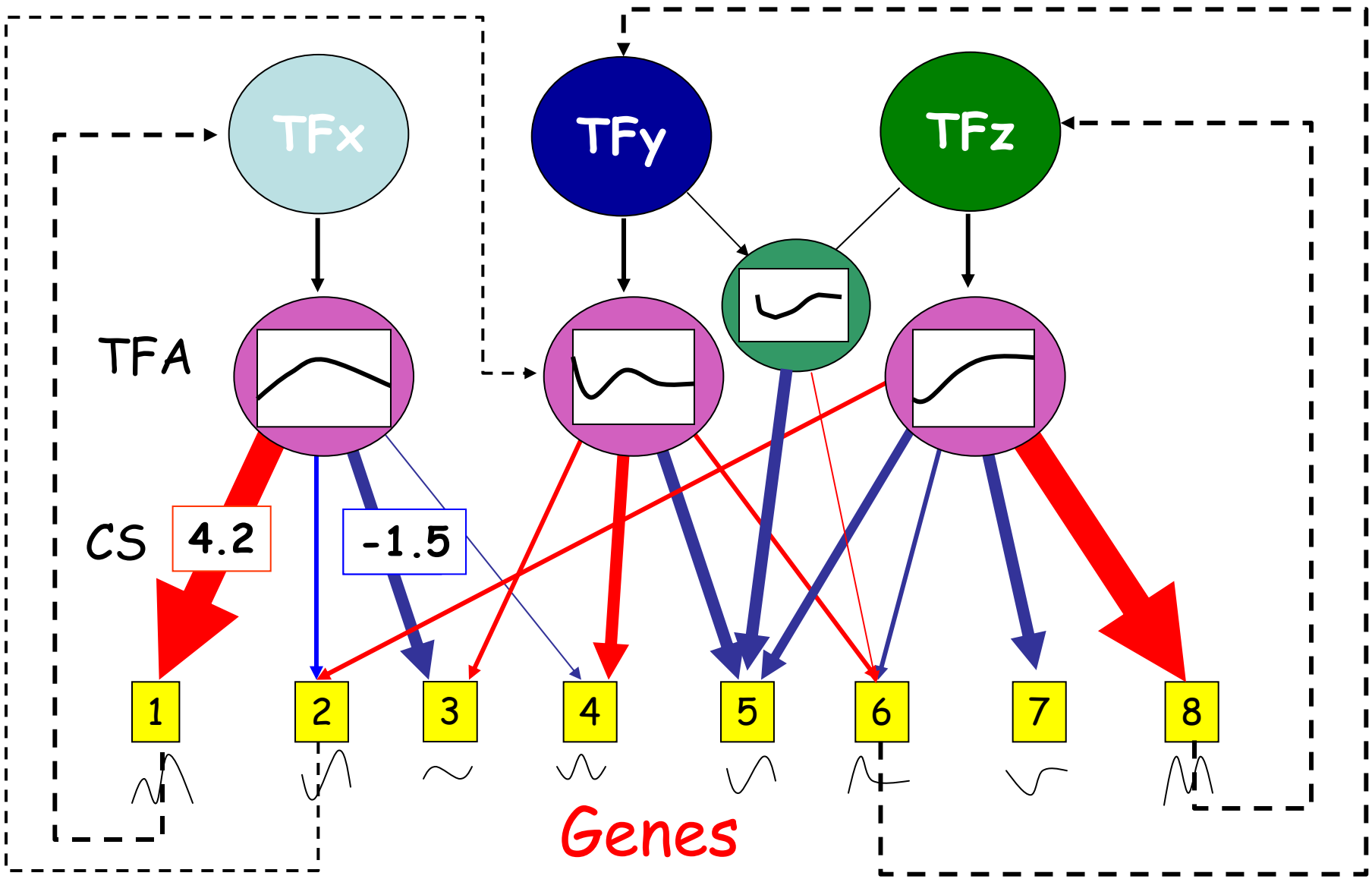
Gene Regulatory Network



Characteristics of Transcription Networks

- Trascription factors regulate transcription
- Transcription factors are regulated post-transcriptionally
- Active transcription factor binds to transcription complex.
- Effect of binding is gene-dependent
- Binding is condition-dependent

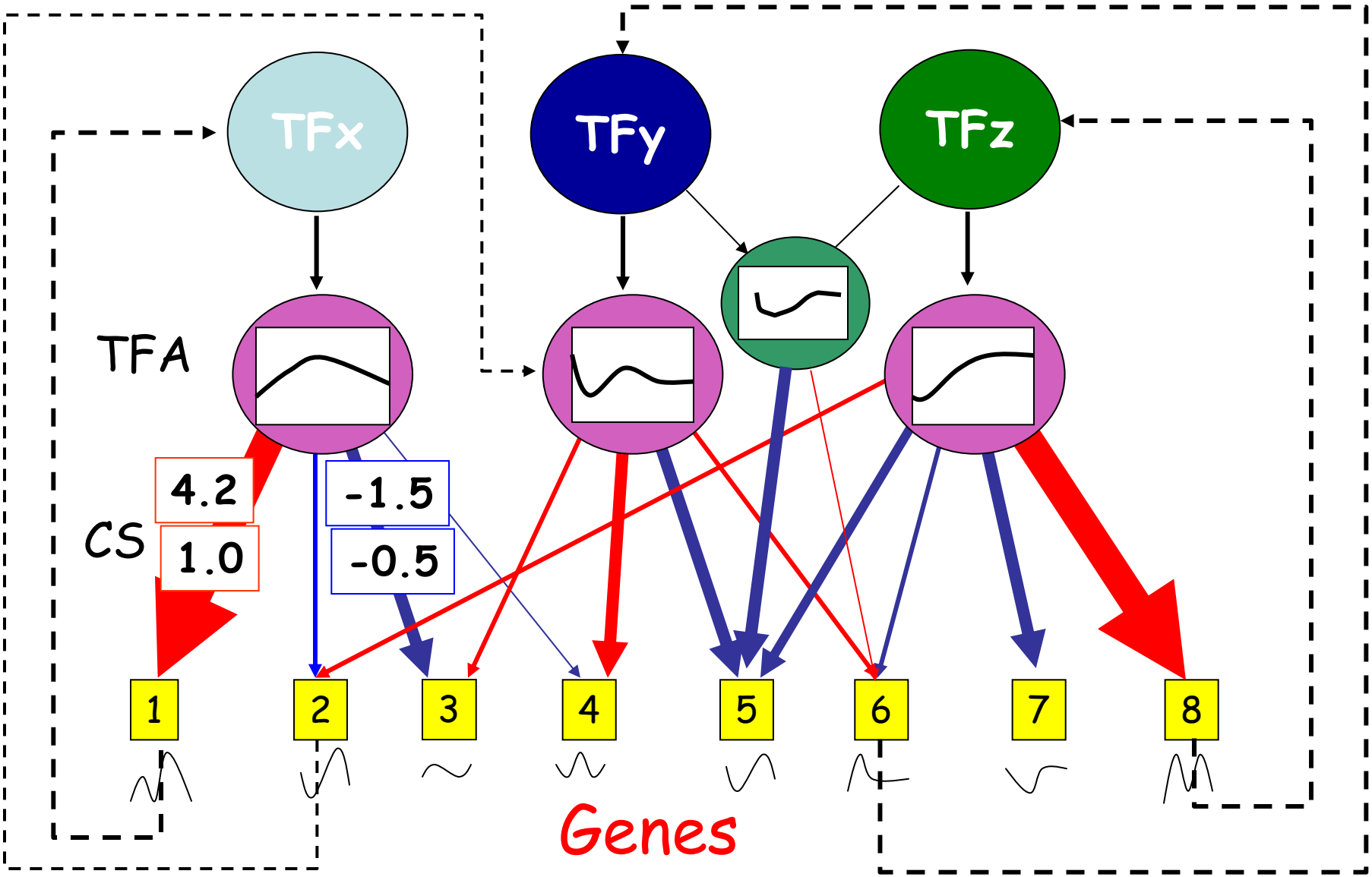
Gene Regulatory Network



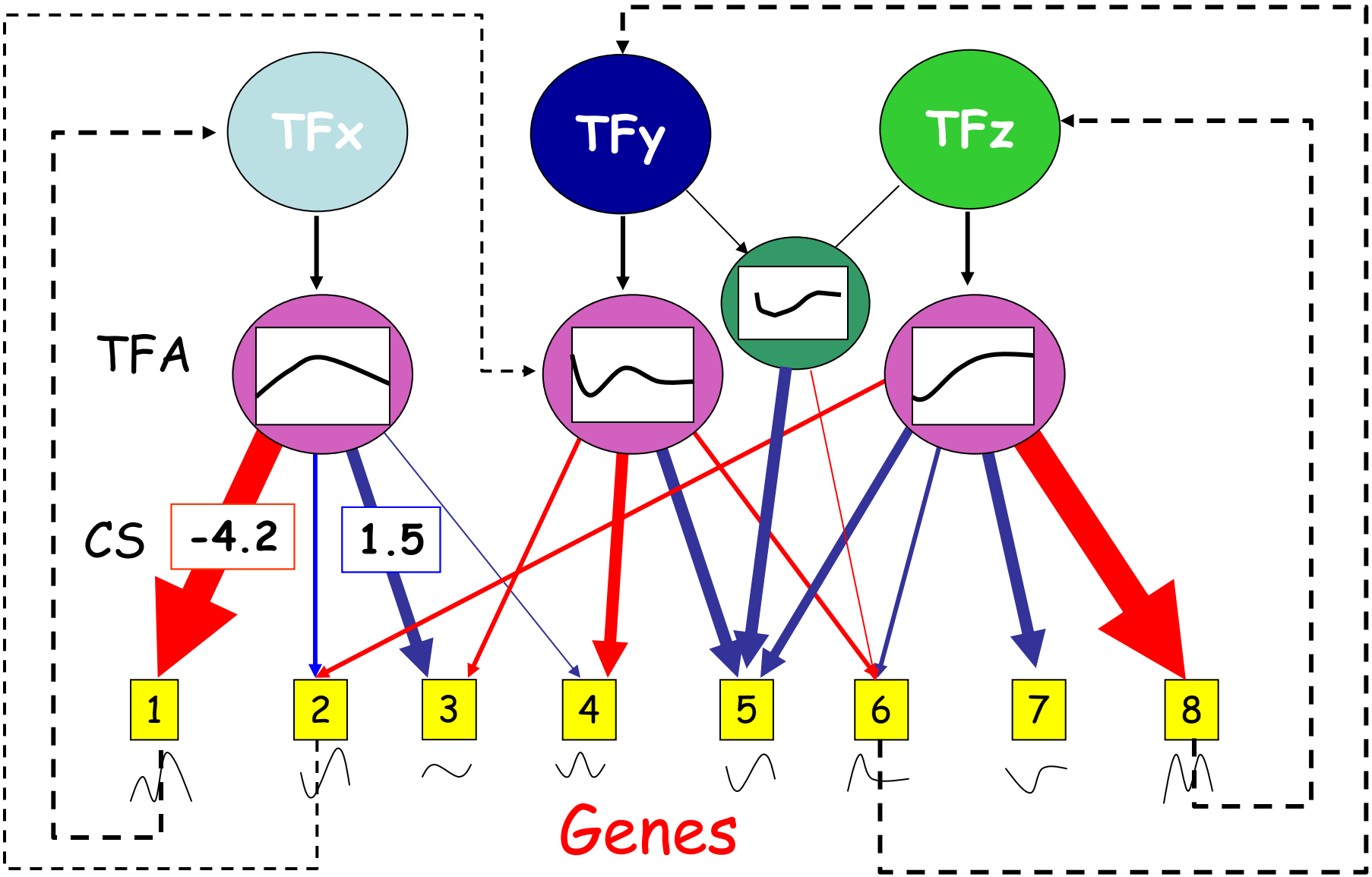
Characteristics of Transcription Networks

- Trascription factors regulate transcription
- Transcription factors are regulated post-transcriptionally
- Active transcription factor binds to transcription complex.
- Effect of binding is gene-dependent
- Binding is condition-dependent

Gene Regulatory Network



Gene Regulatory Network



Characteristics of Transcription Networks

- Trascription factors regulate transcription
- Transcription factors are regulated post-transcriptionally
- Active transcription factor binds to transcription complex.
- Effect of binding is gene-dependent
- Binding is condition-dependent
- A power-law model: not perfect for any gene but roughly good for all genes

Network Component Analysis (NCA)

- Formulation of a mathematical model

$$\frac{d[g_i]}{dt} = V_{synthesis} - V_{\deg radation}$$

$$= \prod TFA_j^{\alpha_{ij}} - k_{di}[g_i] \cong 0$$

Quasi-SS assumption

$$\frac{dTFA_j}{dt} = F_j(g(t), m(t), \theta)$$

Determination of Transcription Factor Activity

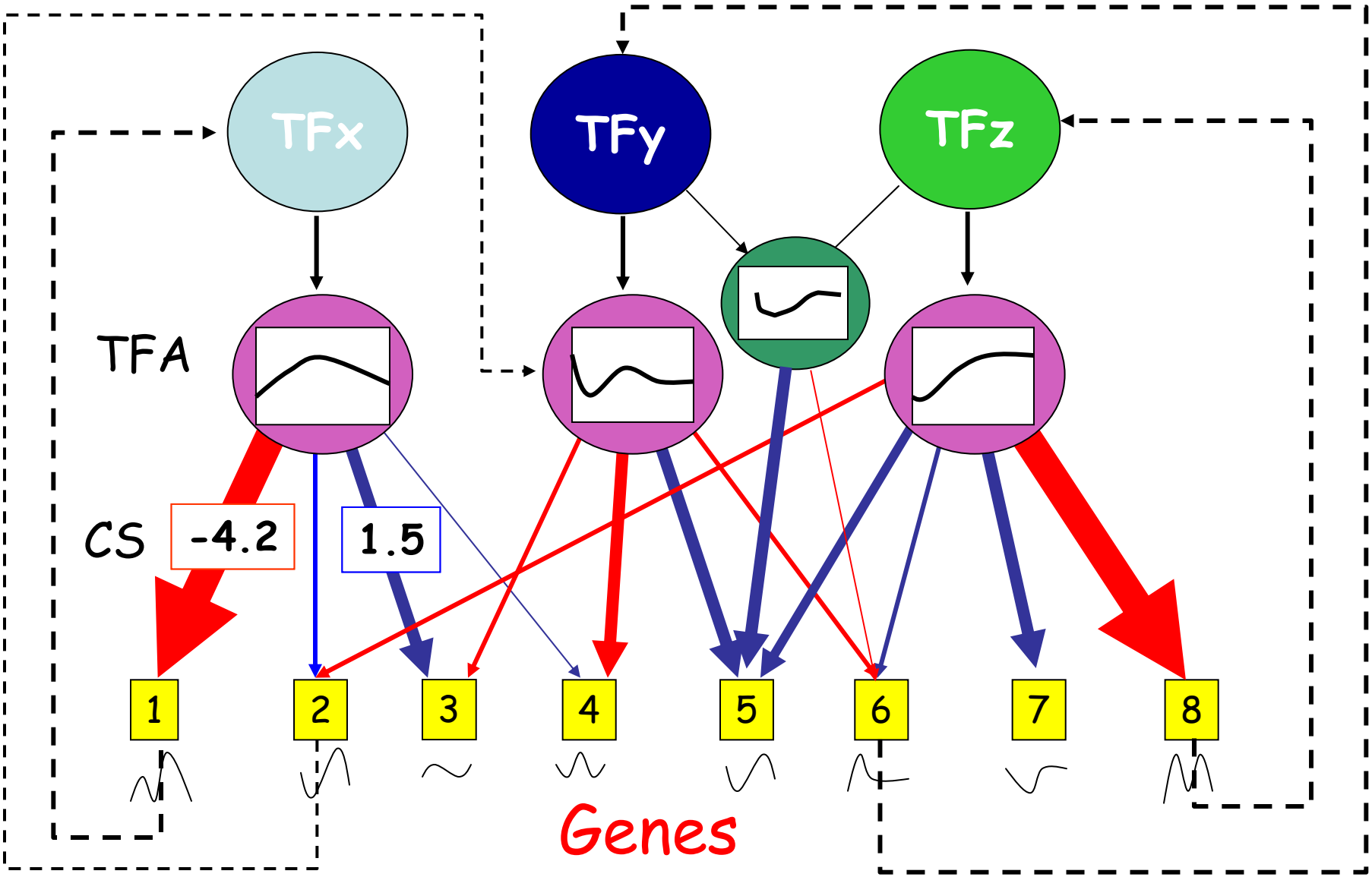
$$\frac{d[mRNA_i]}{dt} = \prod TFA_j^{\alpha_{ij}} - k_{di}[mRNA_i] \cong 0$$

$$\frac{[mRNA_i(t)]}{[mRNA_i(0)]} = \prod \left(\frac{TFA_j(t)}{TFA_j(0)} \right)^{\alpha_{ij}}$$

$$\underbrace{\log\left(\frac{[mRNA_i(t)]}{[mRNA_i(0)]} \right)}_{\text{Microarray data}} = \sum_j \alpha_{ij} \log\left(\frac{TFA_j(t)}{TFA_j(0)} \right) = \underbrace{\sum_j \alpha_{ij} \log\left(\Delta TFA_j(t) \right)}_{\Delta\text{TFA}}$$

$$[E] = [A]\,[P] + \Gamma$$

Gene Regulatory Network



A Mathematical problem.

$$[E] = [A]\,[P] + \Gamma$$

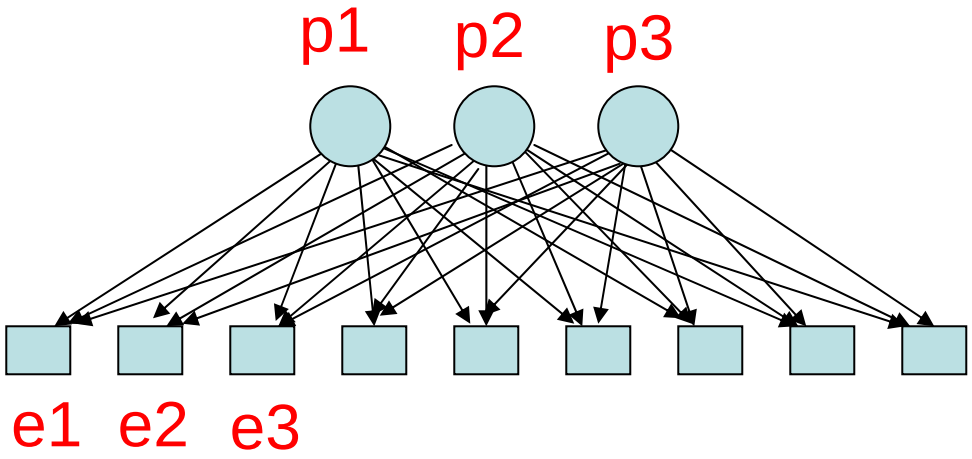
- The matrix decomposition is non-unique.

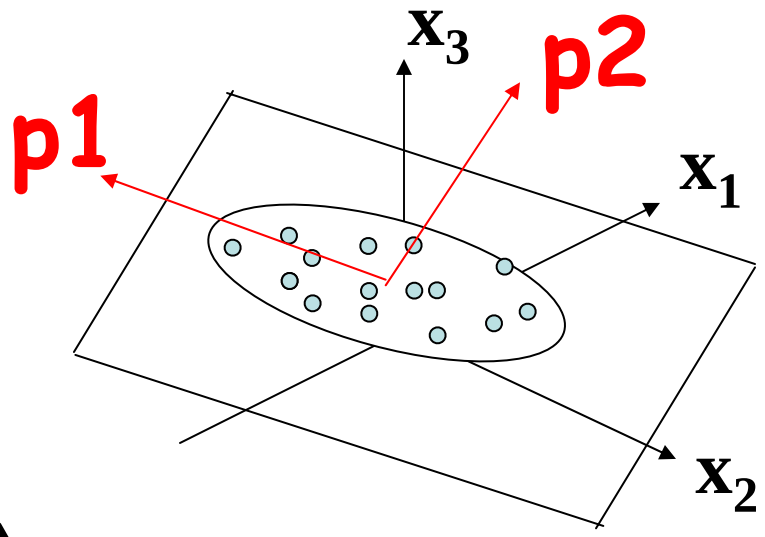
$$[E] = ([A]\,X)\;\;(X^{-1}\,[P]) + \Gamma$$

- Unless it is properly constrained.

Principal Component Analysis (PCA or SVD)

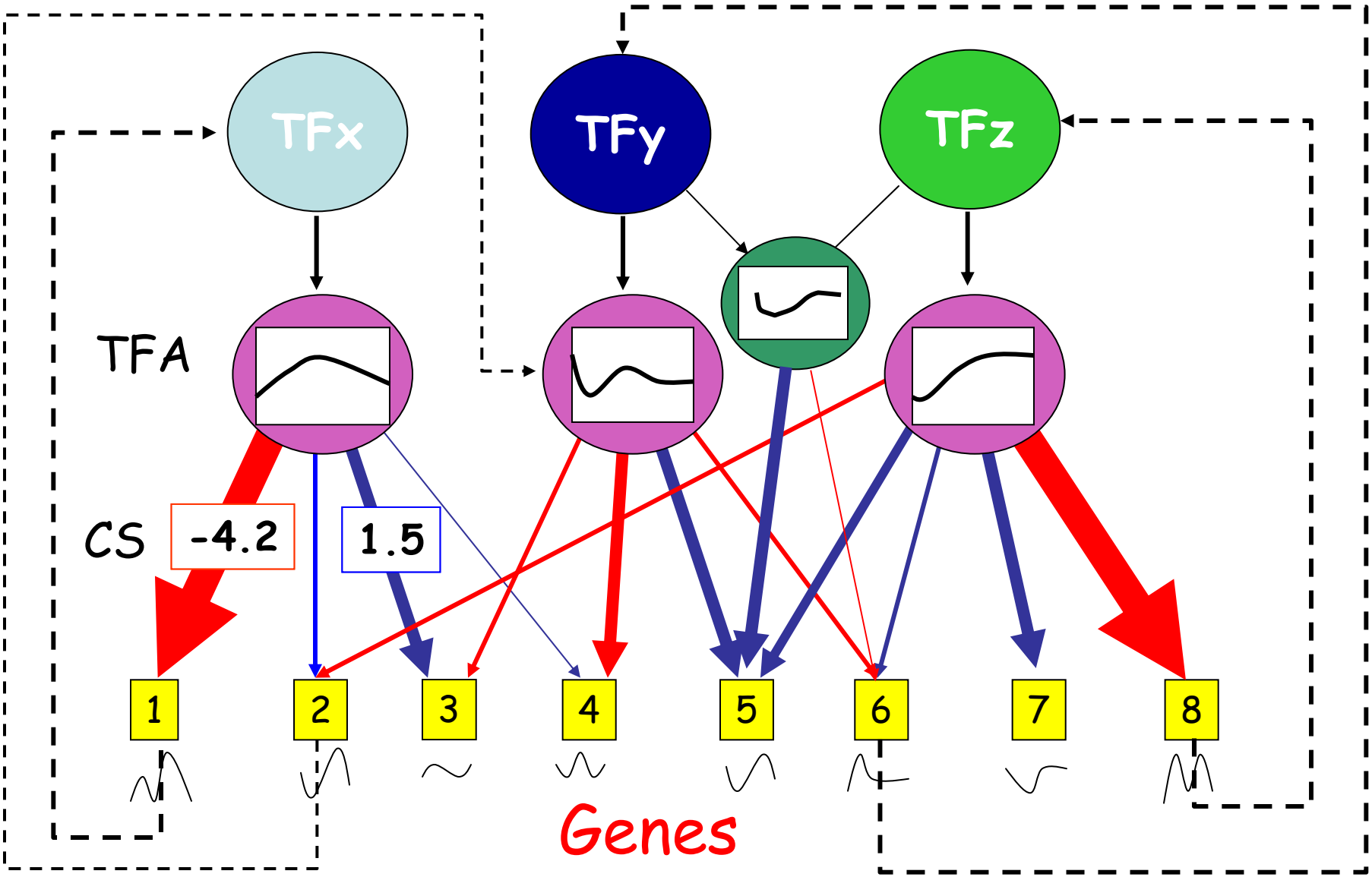
$$[E]_{NxM} = [A]_{NxL}[P]_{LxM}$$





- Explains major variations in data
- Useful for visualization and classification
- No mechanistic insight

Gene Regulatory Network



Modeling

- What is the purpose of the model?
  - To elucidate the NO response network.
- What are the existing/alternative approaches?
  - ad hoc analysis, clustering…
- What is the scope of the model?
  - mRNA, transcription factors, degradation factors…
- What data to use?
  - Microarray transcriptome data
- Mathematical details?
- What is the basis of prediction?

Network connectivity provides constraints

$$[E] = [A]\,[P] + \Gamma$$

Constraints: $a_{ij} = 0$

$$for \quad a_{ij} \in Z_A$$

$$\begin{bmatrix} * & 0 & * & 0 & * \\ 0 & * & 0 & * & 0 \\ 0 & 0 & * & * & 0 \\ * & * & 0 & 0 & * \\ * & 0 & * & 0 & 0 \end{bmatrix}$$

Network Component Analysis (NCA)

$$[E]_{NxM} = [A]_{NxL}[P]_{LxM} \quad \text{s.t.} \quad a_{ij} \in Z_A, \; p_{ij} \in Z_P$$

If

1. Each column of A has at least L-1 zeros
2. The non-zero members of a column should not be a subset of others.
3. The sub-matrix P involved in each gene has full row rank.

then, the decomposition is unique up to a scaling matrix.

NCA forms a bi-linear optimization problem

$$E = AP + \Gamma$$

$$\min_{A,P} \| E - AP \|^2$$

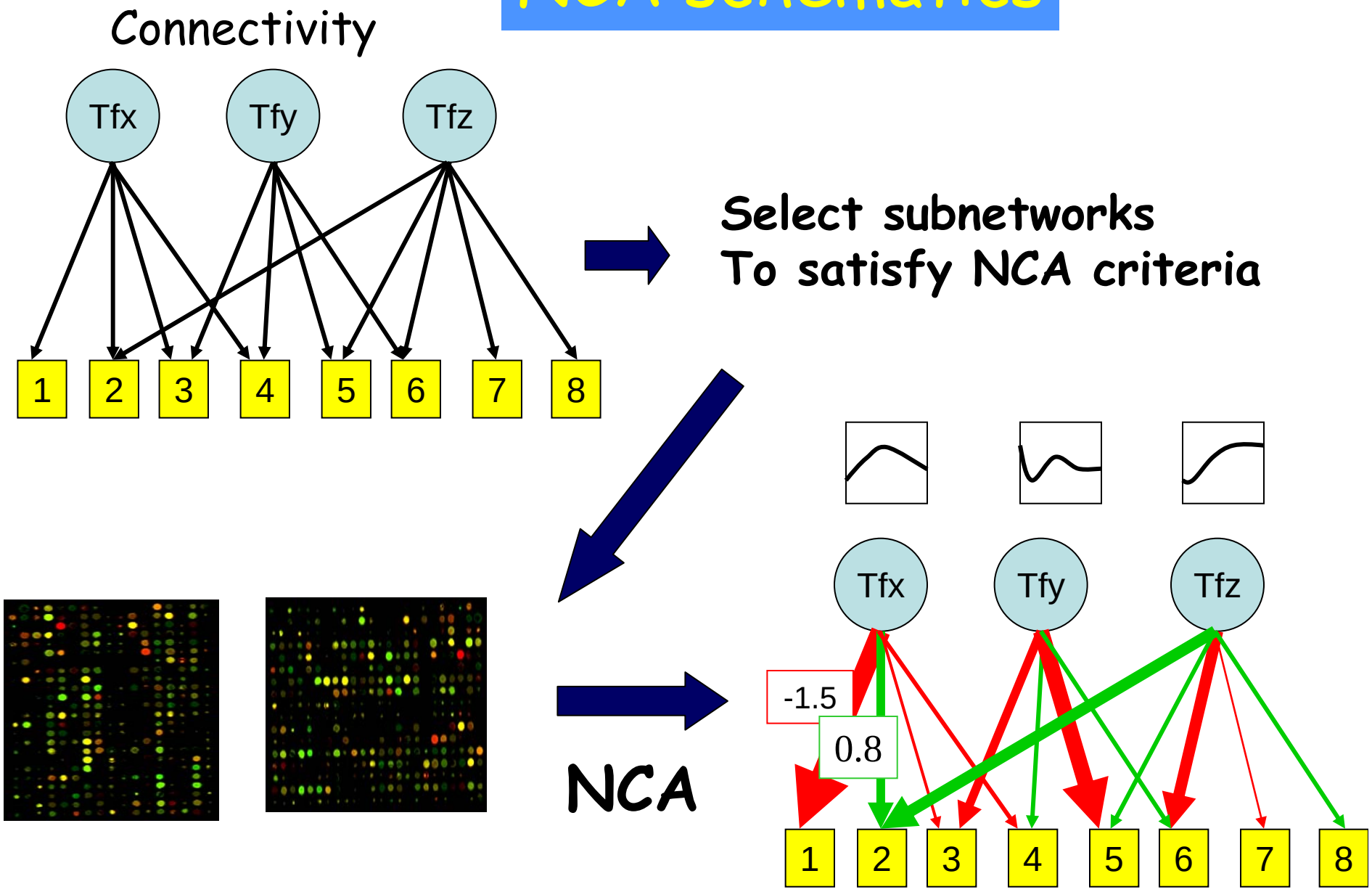
$$\text{s.t. } A \in Z_A \quad \leftarrow \text{Connectivity constraints}$$

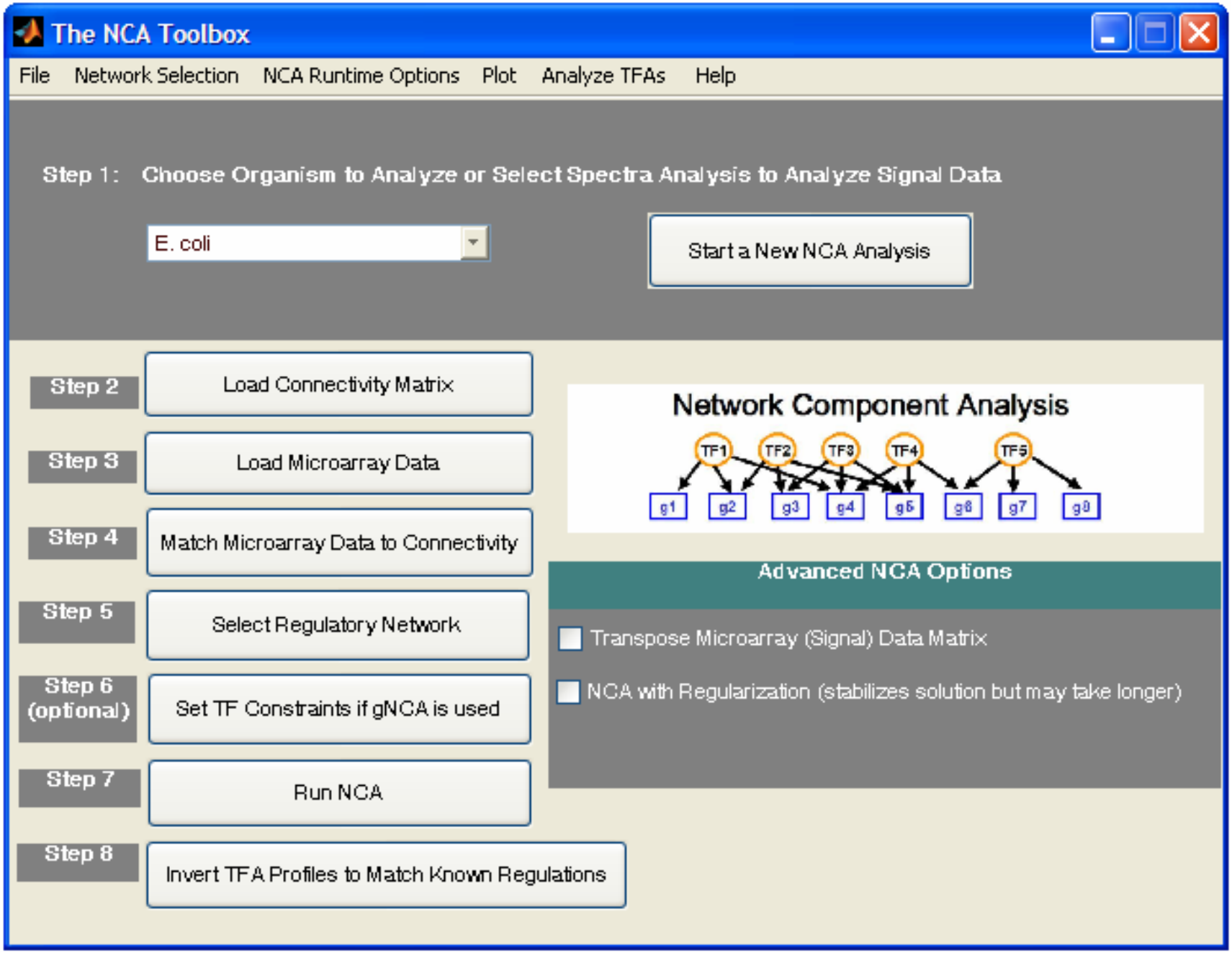
$$P \in Z_P \quad \leftarrow \text{Regulatory constraints}$$

Network Component Analysis (NCA)

- Biological model
  - Gives biological significance and limitation
- Identifiability criteria
  - Yield mathematical constraints need to be satisfied
- Connectivity theorems
  - Generate insight into connectivity density

NCA schematics




The NCA Toolbox
File
Network Selection
NCA Runtime Options
Plot
Analyze TFAs
Help
Step 1: Choose Organism to Analyze or Select Spectra Analysis to Analyze Signal Data
E. coli
Start a New NCA Analysis
Step 2
Load Connectivity Matrix
Step 3
Load Microarray Data
Step 4
Match Microarray Data to Connectivity
Step 5
Select Regulatory Network
Step 6 (optional)
Set TF Constraints if gNCA is used
Step 7
Run NCA
Step 8
Invert TFA Profiles to Match Known Regulations
Network Component Analysis
TF1
TF2
TF3
TF4
TF5
g1
g2
g3
g4
g5
g6
g7
g8
Advanced NCA Options
Transpose Microarray (Signal) Data Matrix
NCA with Regularization (stabilizes solution but may take longer)

Modeling

- What is the purpose of the model?
  - To elucidate the NO response network.
- What are the existing/alternative approaches?
  - ad hoc analysis, clustering…
- What is the scope of the model?
  - mRNA, transcription factors, degradation factors…
- What data to use?
  - Microarray transcriptome data
- Mathematical details?
- What is the basis of prediction?

Basis for Prediction

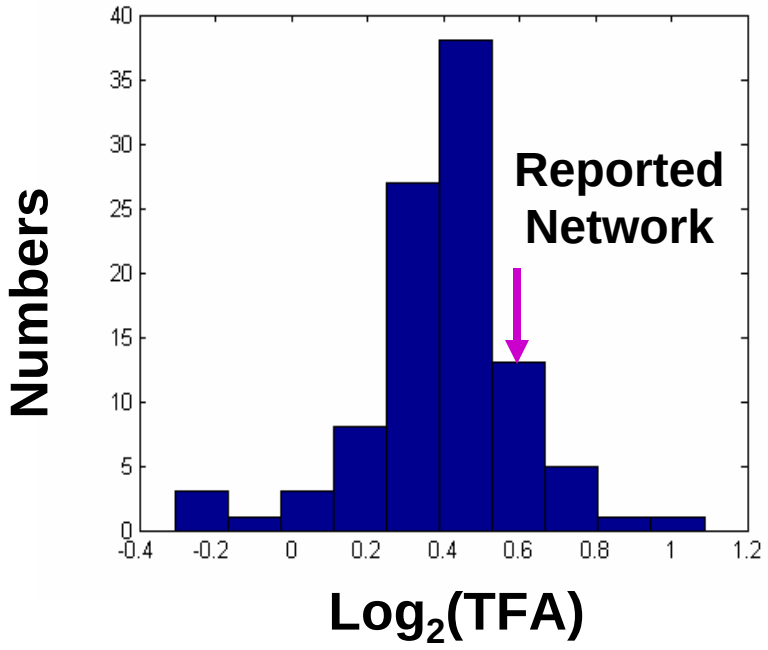
- Transcriptome data (mRNA abundance)
- Transscription connectivity
- Power-law model

Active transcription factors
Active networks

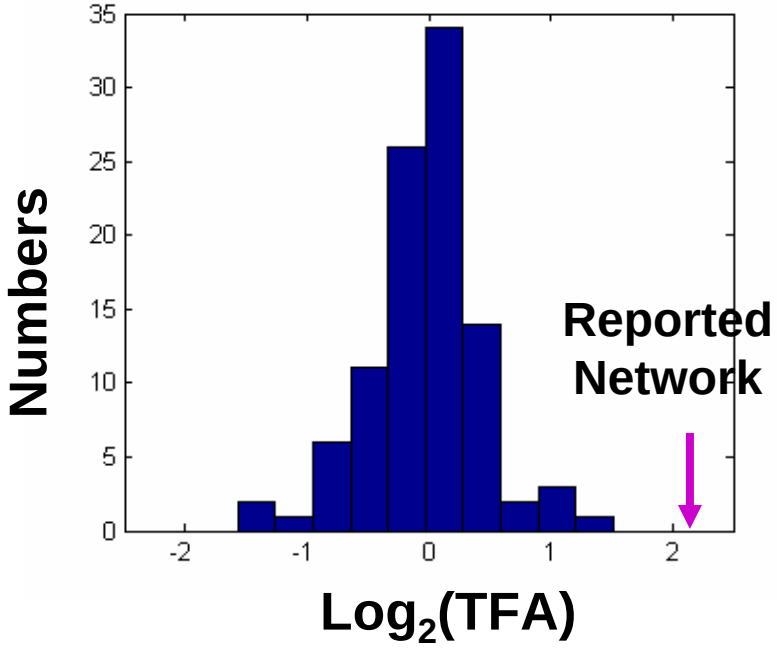
Statistical Significance of TFA

- Constructing null distribution of TFA by *random networks*
  - ✓ Expression data: randomly sampling from genenome
  - ✓ Connectivity: the same with reported network

Null distribution of E2F TFA in Akt-Tg after inducing 12hrs

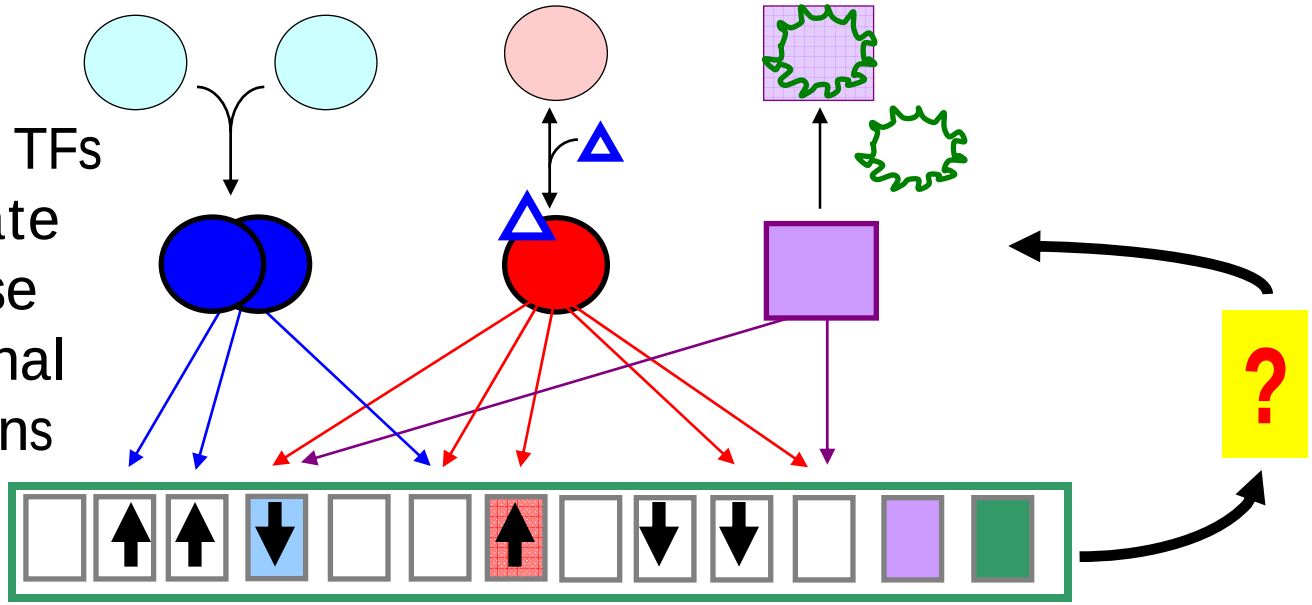


Null distribution of MYBL2 TFA in Pten-null prostate at PIN state



NCA Identifies NO Response Regulators

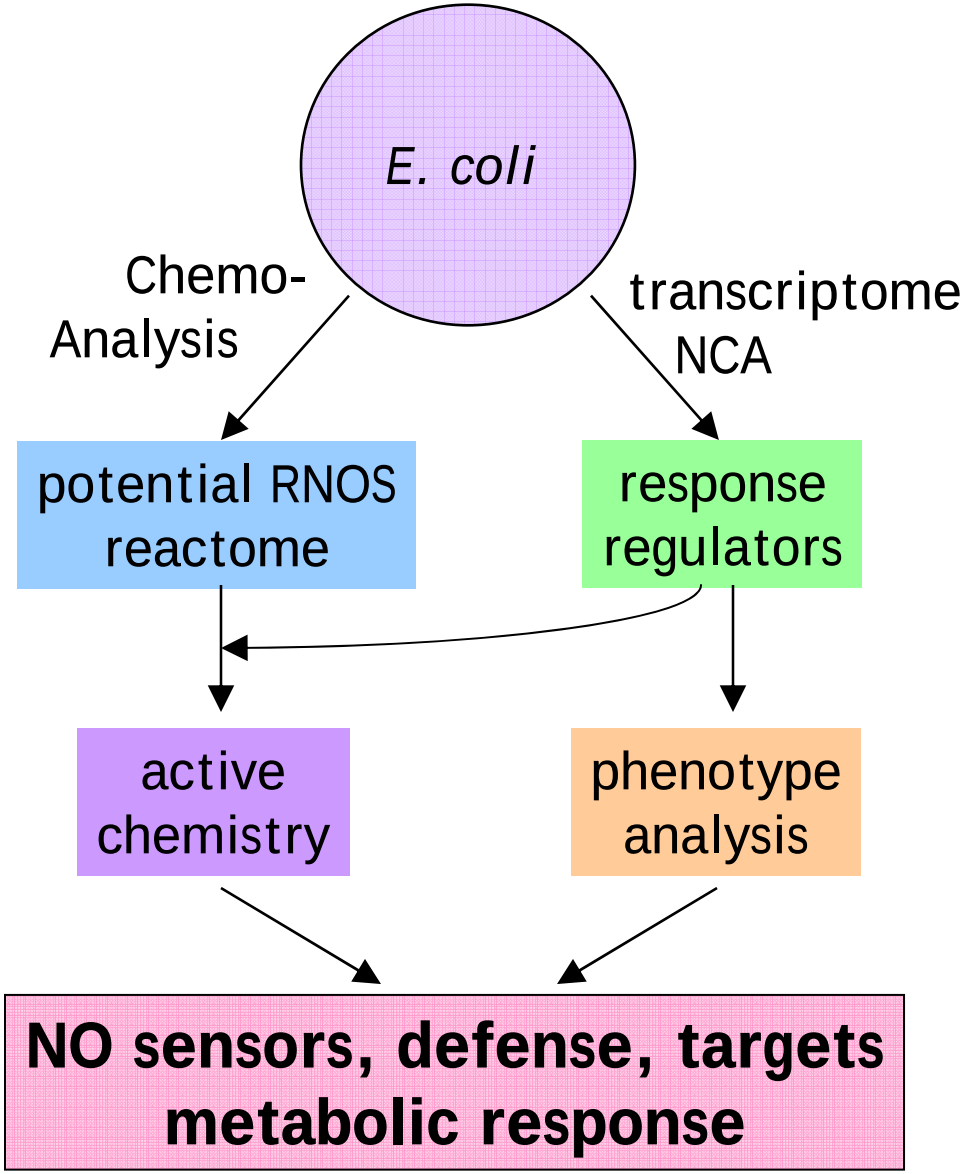
Goal: Identify TFs which mediate NO response transcriptional perturbations

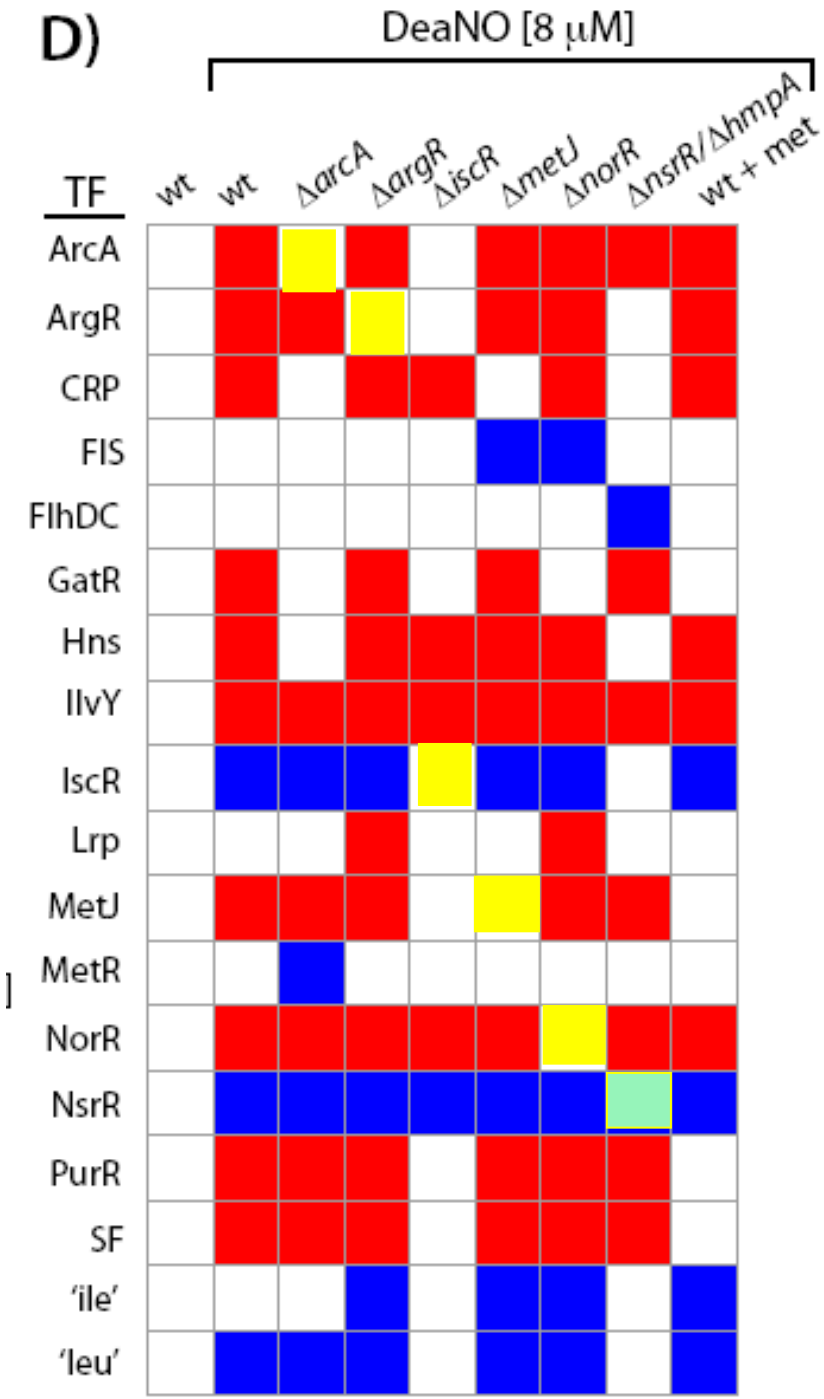


use DNA microarray to measure network behavior

AcrR	Ada	**ArcA**	**ArgR**	ArsR	AsnC	AtoC	BetI	BirA	CadC	CbI
CpxR	**Crp**	CspA	CsrA	CynR	CysB	DcuR	EmrR	ExuR	FadR	FecI
FhlA	Fis	FlhDC	FliA	Fnr	FruR	Fur	**GatR**	HcaR	HipB	his
HNS	IclR	IdnR	IHF	ile	**IlvY**	**IscR**	KdpE	LexA	**leu**	Lrp
LysR	MalT	MaoB	**MetJ**	MetR	MhpR	Mlc	ModE	NadR	NagC	NarL
NorR	**NsrR**	OxyR	PhoB	phe	**PurR**	RcaSB	RhaSR	Rob	RpiR	RpoE
RpoH	RpoN	RpoS	RtcR	SoxR	trp	TyrR	XapR	XylR	YjbK	**SF**

Chemogenomic Approach





NCA-derived TF activity change

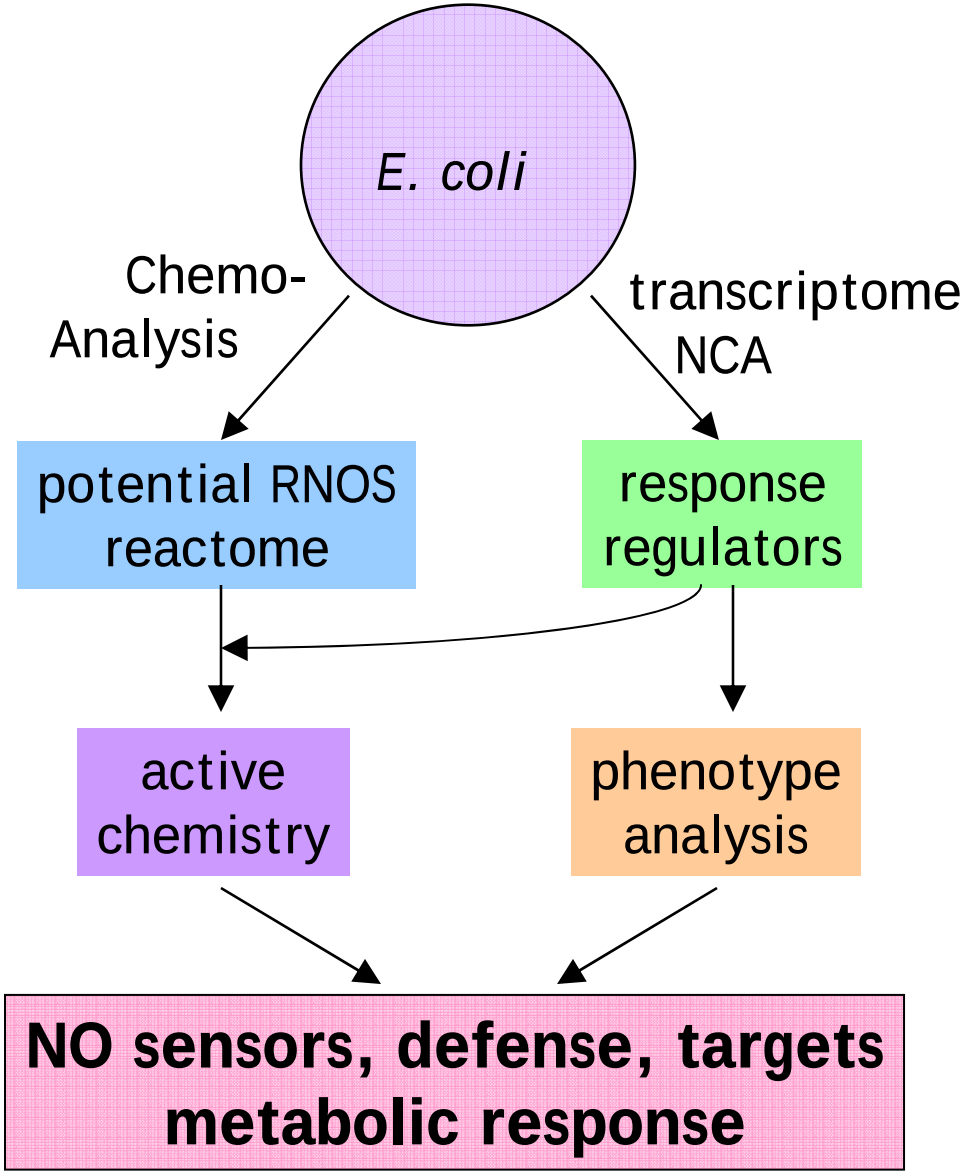
Increased Activity

Decreased Activity

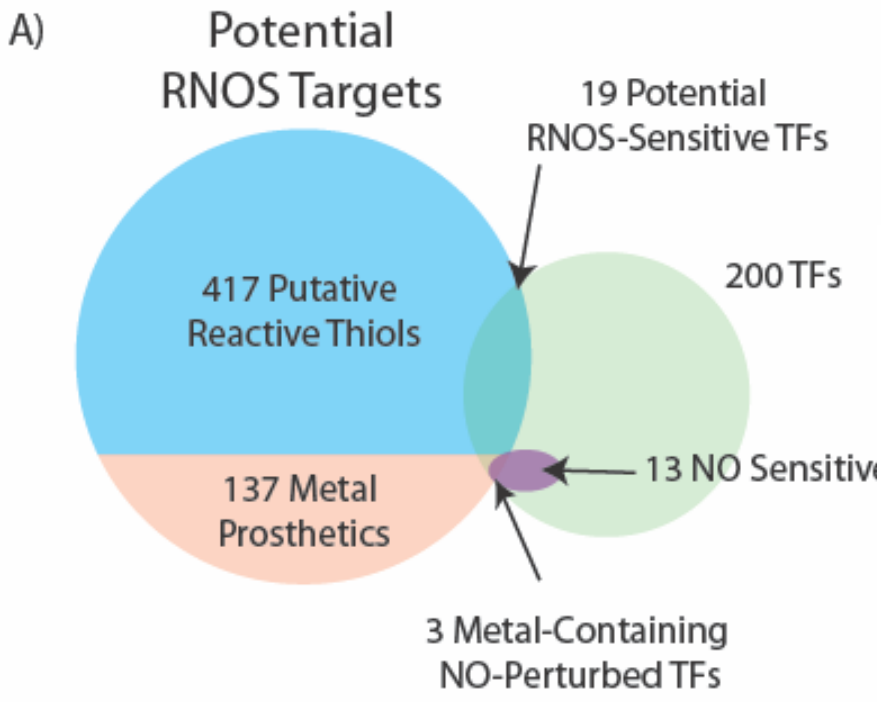
TF deletion confirming expected null activity

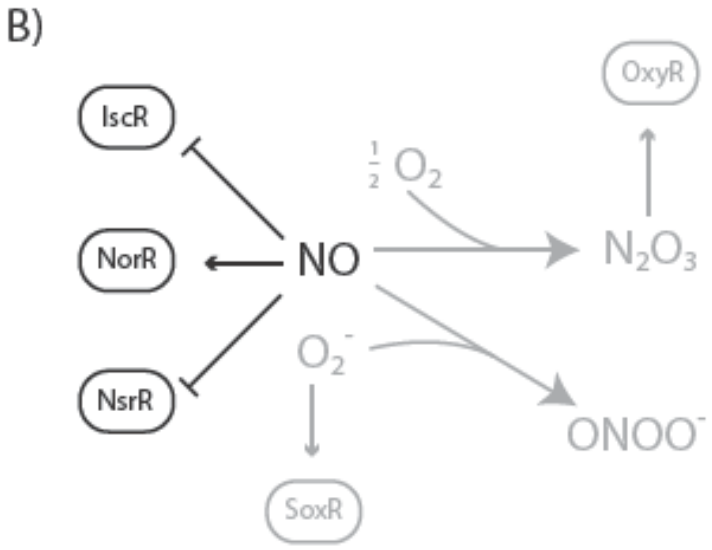
TF deletion confirming expected activity (decreased)

Chemogenomic Approach



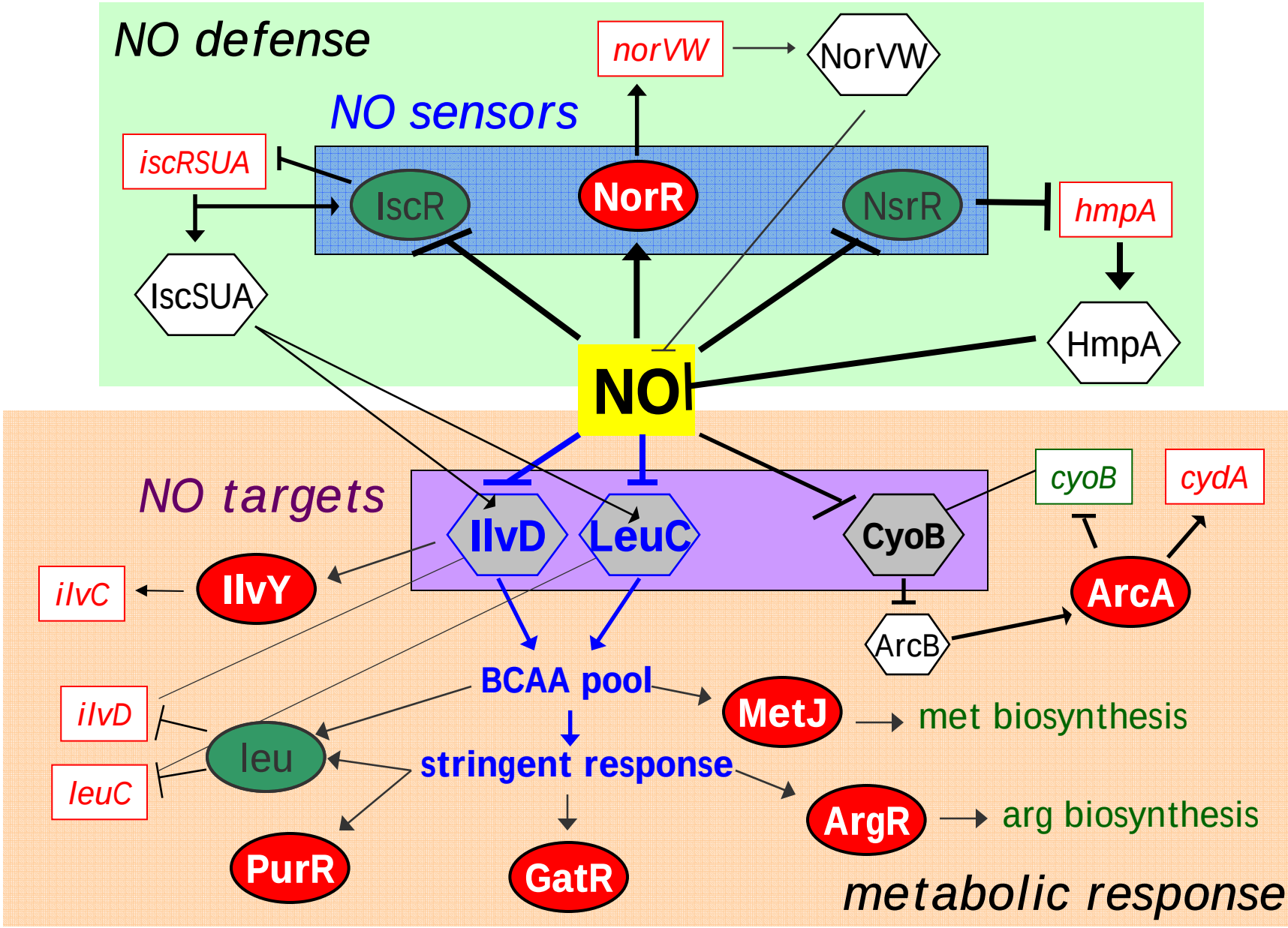
Combination of Active TFs and Reactome Identifies chemistry involved.



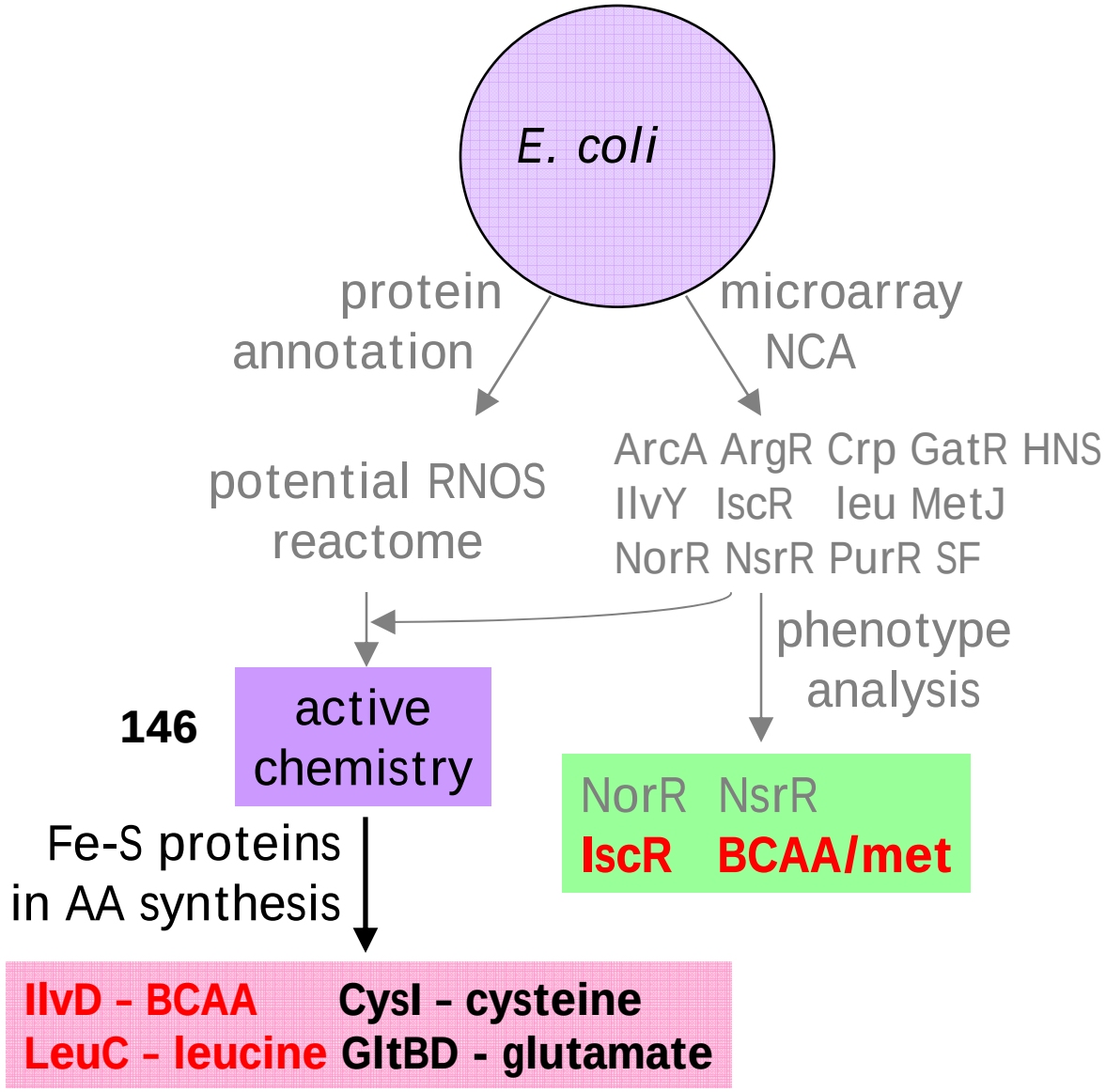


Direct NO-metal interaction is the chemistry involved,

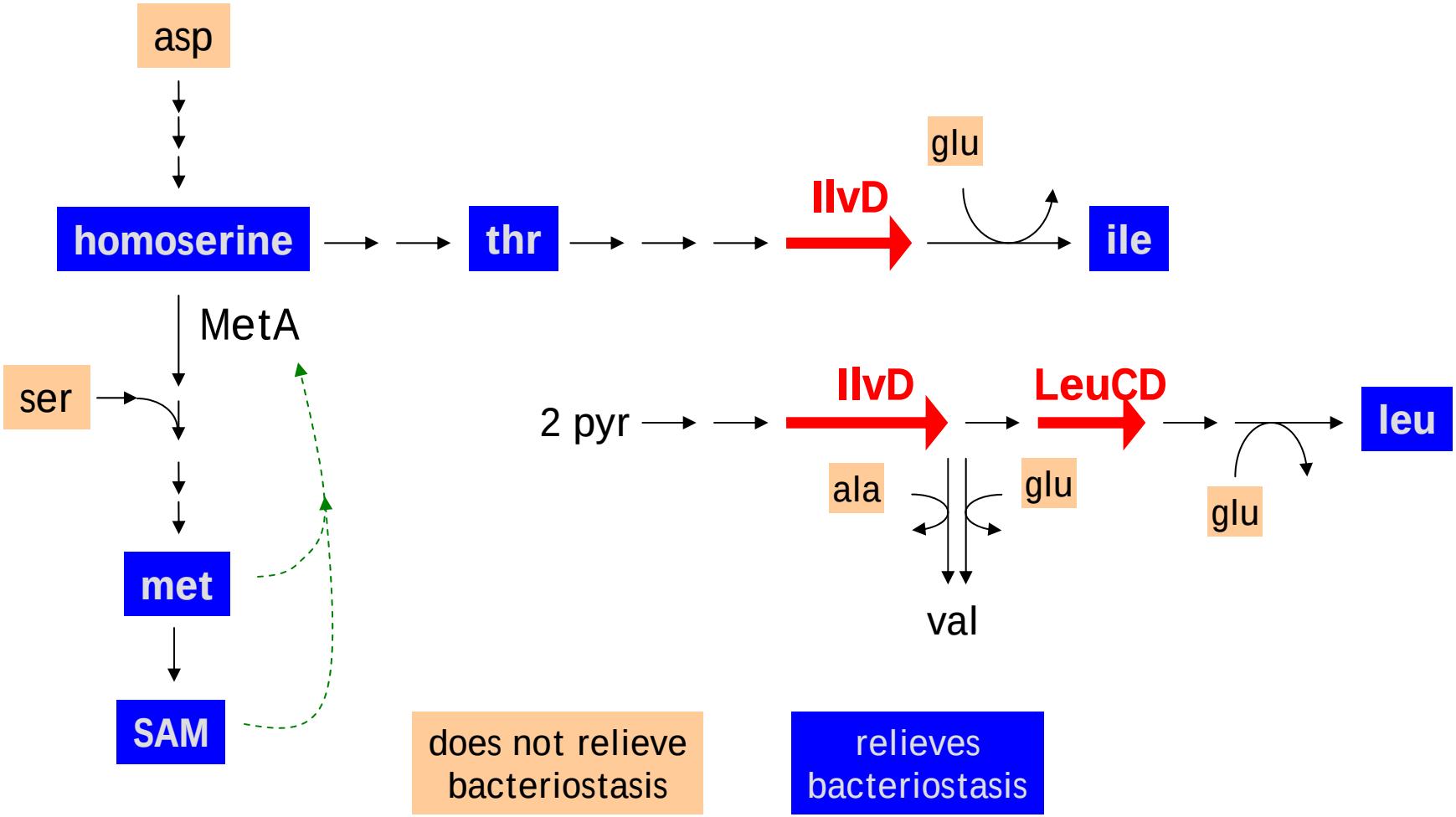
NO Response Network



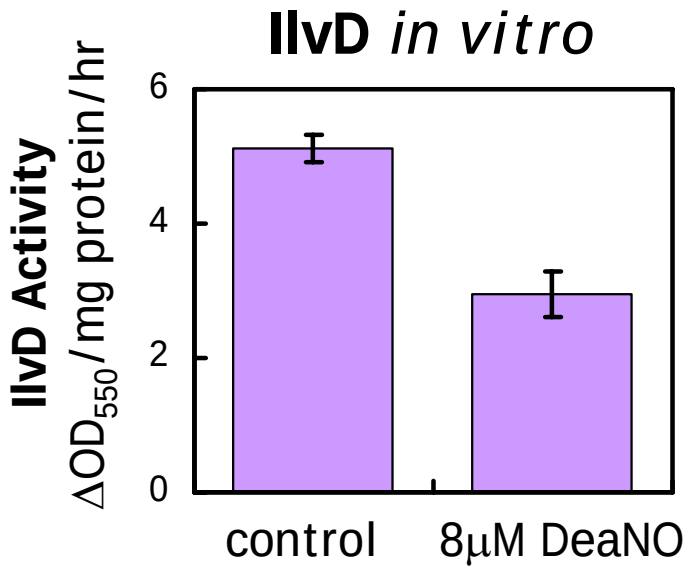
Fe-S Proteins in AA Synthesis

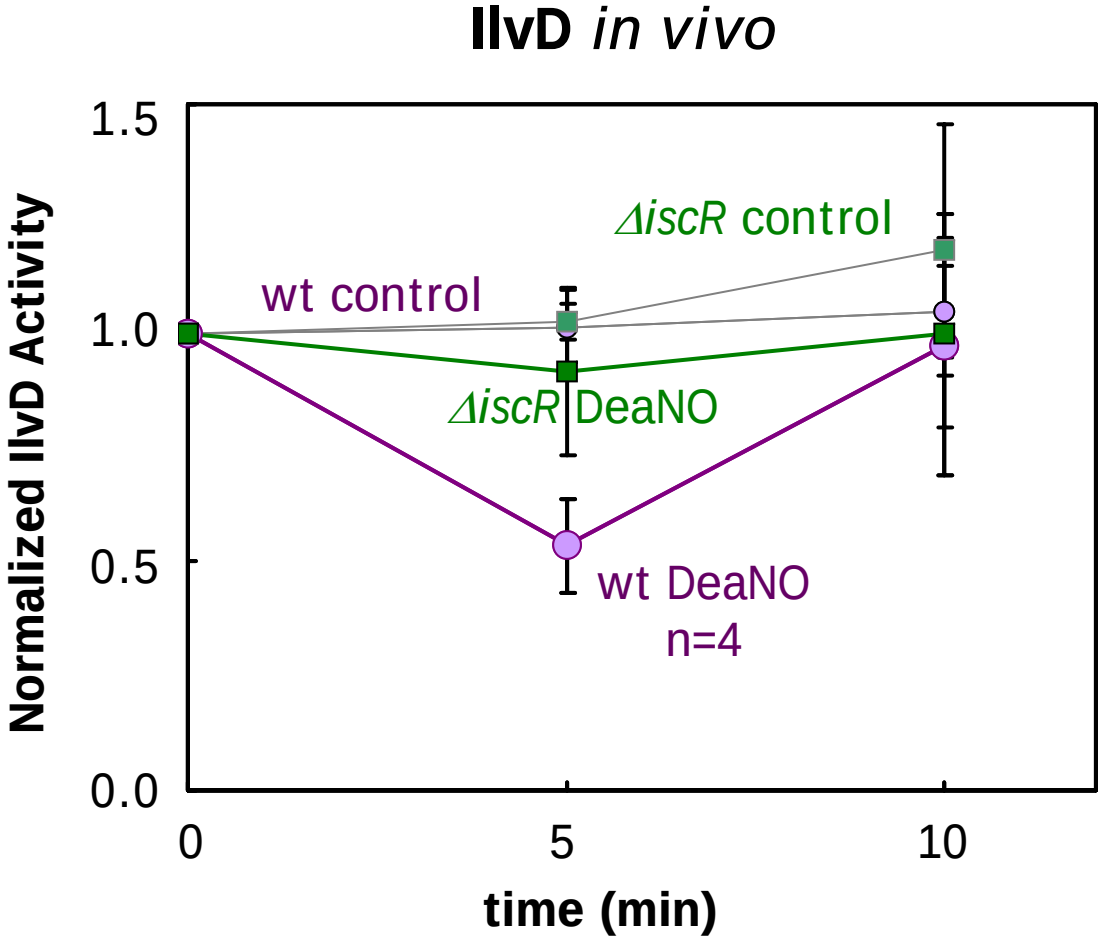


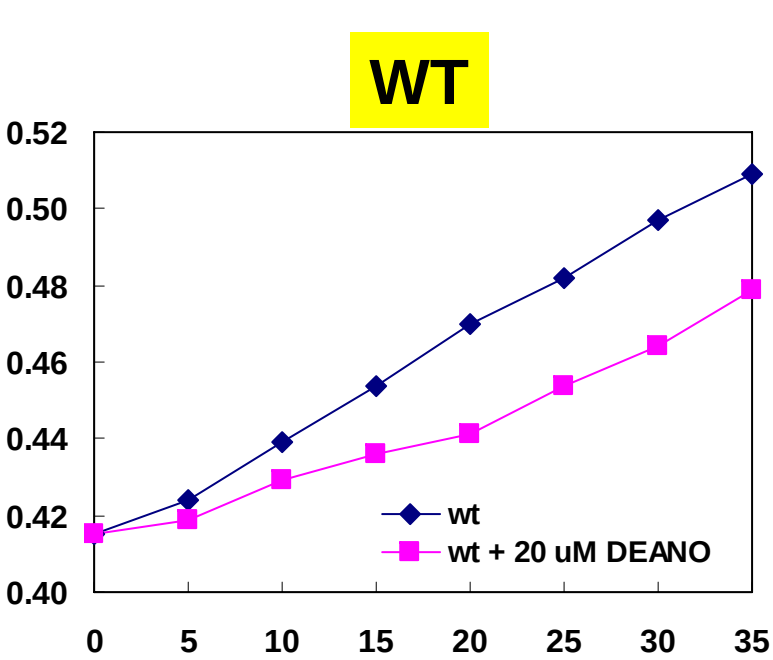
BCAA/met Pathway

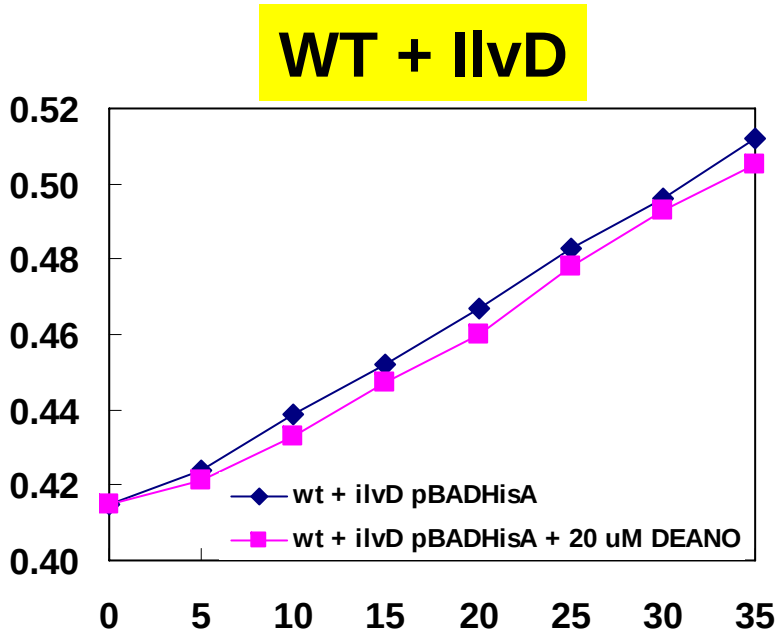


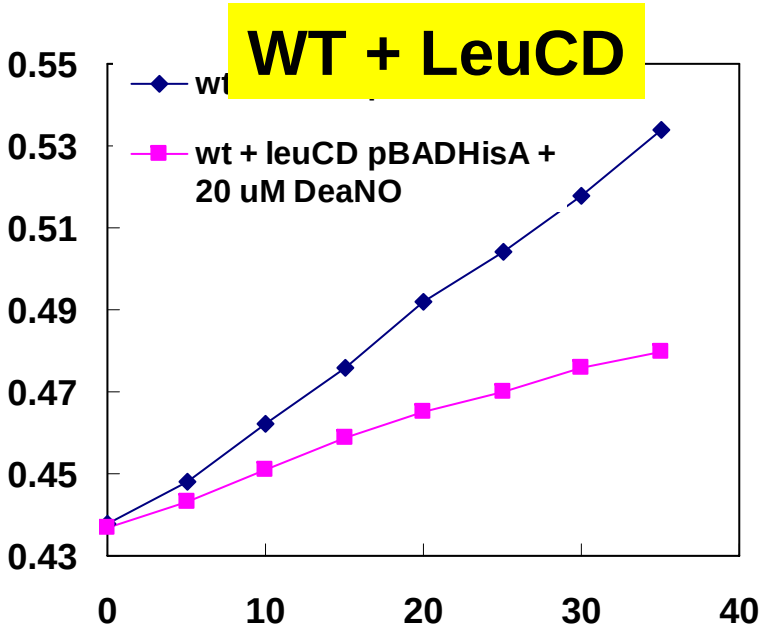
IlvD is NO Targets



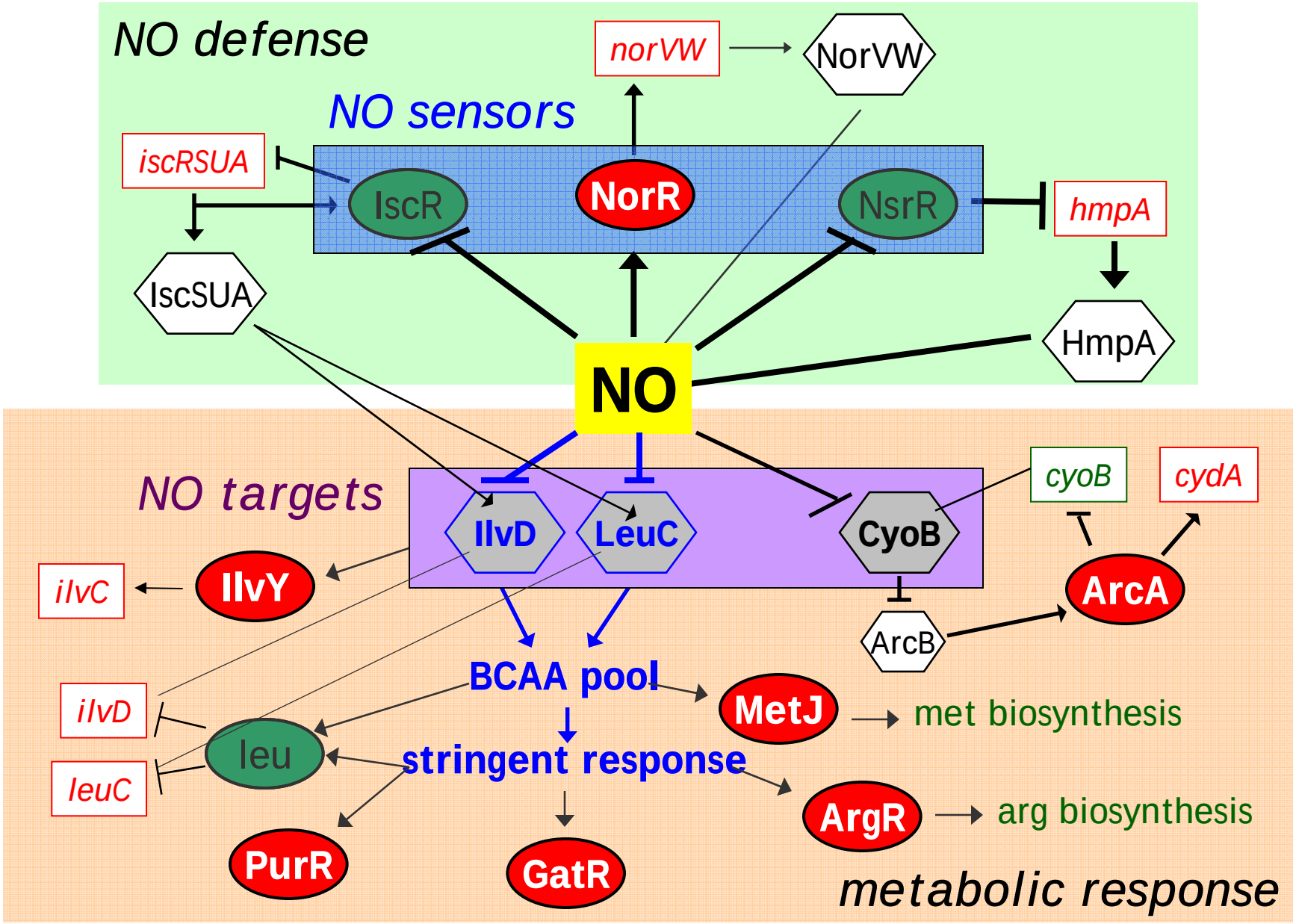



WT
0.52
0.50
0.48
0.46
0.44
0.42
0.40
0
5
10
15
20
25
30
35
wt
wt + 20 uM DEANO


WT + IlvD
0.52
0.50
0.48
0.46
0.44
0.42
0.40
0
5
10
15
20
25
30
35
wt + ilvD pBADHisA
wt + ilvD pBADHisA + 20 uM DEANO


WT + LeuCD
0.55
0.53
0.51
0.49
0.47
0.45
0.43
0
10
20
30
40
wt + leuCD pBADHisA +
20 uM DeaNO

NO Response Network



Acknowledgment